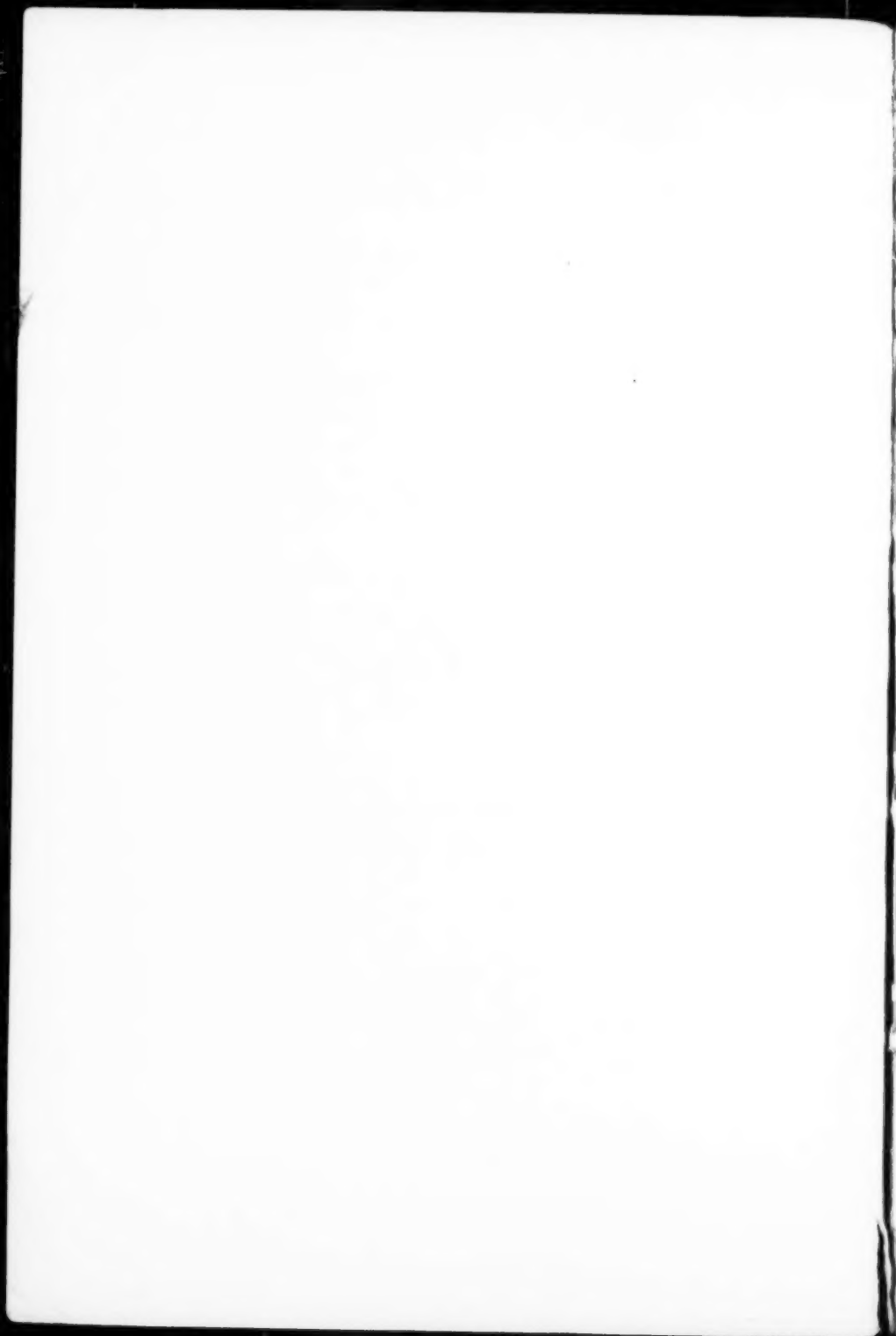# NATIONAL
CONVENTION
TRANSACTIONS
1958

TWELFTH ANNUAL CONVENTION

AMERICAN SOCIETY
FOR
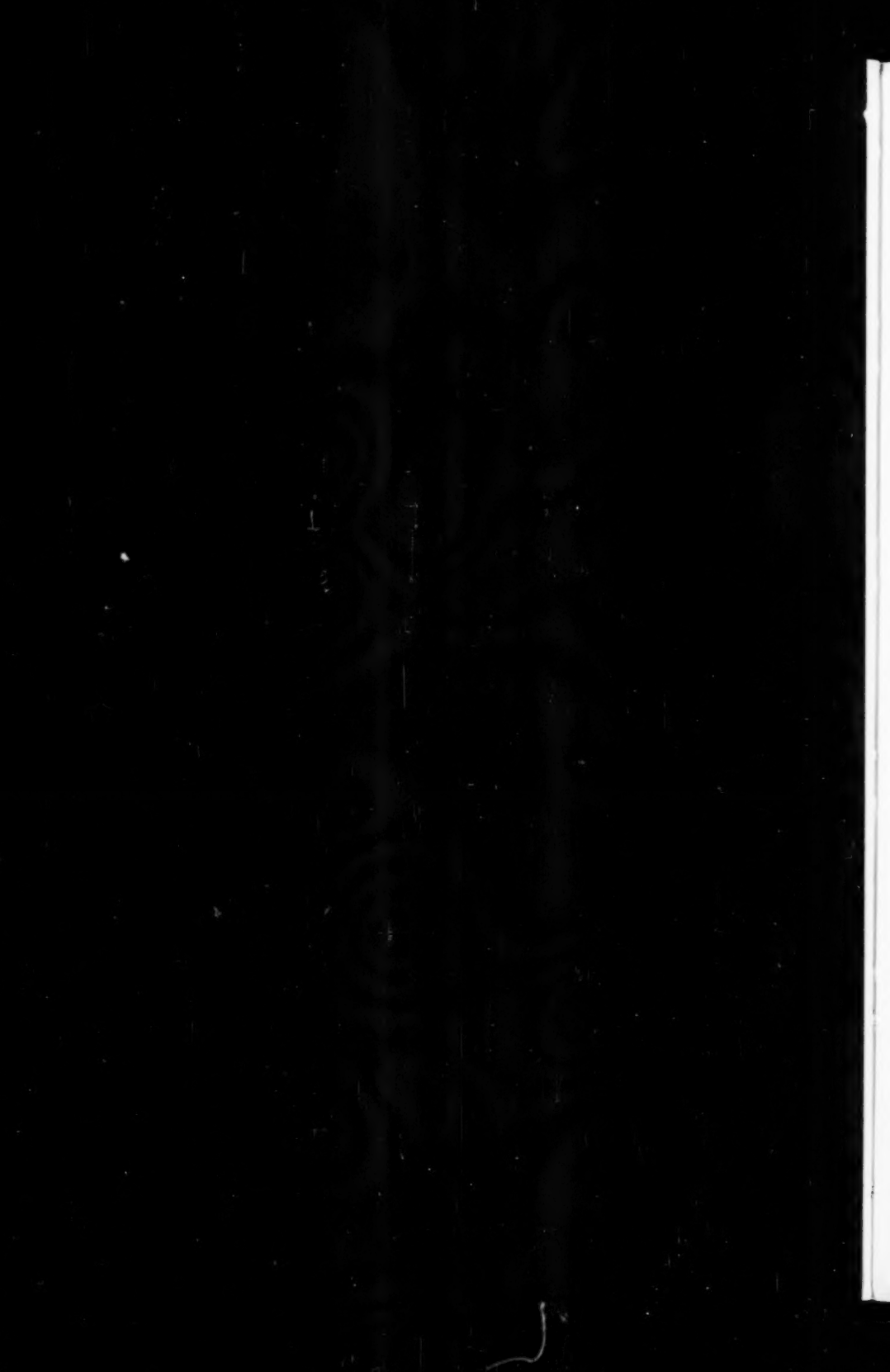# QUALITY CONTROL

NATIONAL
CONVENTION
TRANSACTIONS
1958

TWELFTH ANNUAL CONVENTION

AMERICAN SOCIETY FOR QUALITY CONTROL

MAY 26, 27, 28, 1958

BOSTON, MASSACHUSETTS

PRICE
$3.00 at the Convention
$3.50 if Ordered by Mail

EFFECTIVE QUALITY CONTROL PROGRAM FOR THE INDUSTRIAL CONTROL LABORATORY

Frank W. Kroll, Jr.
Esso Standard Oil Company

Esso Standard's Bayway Refinery has pioneered in the effective appli-
cation of statistical quality control to improve the management of the
control laboratory. While this application is only the first phase of a
broad program, the reader will find a well integrated laboratory quality
control program described in this paper. The net result of making the
"quality control function" an integral part of the normal line-super-
visory responsibility has been improved testing reliability for many
standard, well established chemical and physical test methods.

Introduction

An oil refinery produces a large number of products and performs
intermediate and finished-product inspections by the thousands every day.
To do this inspection job, many tests are employed by staffs of trained
laboratory technicians. The measurement of product quality has become
increasingly important in the post-war years along with an increasingly
competitive market for petroleum-refined products. Lack of reliable
product quality measurements, in this kind of situation, may be respon-
sible for excessive "product-quality giveaway" and improvement in the
accuracy and precision of test methods leads to real dollar savings to a
refinery. It was primarily for this reason that, two years ago, organi-
zation and direction was given to the use of statistical quality control
as a management tool to improve the reliability of laboratory inspection
results. Alternate ways of accomplishing our objective were considered.
Multi-testing (analyzing enough replicate tests to reduce the standard
deviation of the reported result), development of new test methods,
increased emphasis on laboratory technician training, revision and
simplification of tests and the rewriting of test procedures were all
tried; none of these approaches accomplished the job of improving and
maintaining work-quality effectively.

It is the purpose of this paper to describe the overall laboratory
quality control program and to point to some of the important details
which have a bearing on the success of such a program.

Management Philosophy of the Program

The single most important factor which prevents a quality control
program from becoming a one-shot affair (usually in the form of a
committee activity) is the continuing interest and patronage of top-
management. To accomplish this, management must be "inspired" by a
purposeful and practical philosophy. For the industrial inspection
laboratory, this philosophy must, of necessity, involve all personnel
concerened with the production of the "laboratory commodity - the
inspection result". It must involve, in particular, those who dispense
this commodity to the customers - the line-supervisory foreman. It is
axiomatic that continued acceptance of the commodity by customers
depends on the continued high-quality of the daily routine inspection
result. With these concepts in mind, it was decided to fortify the
laboratory foremen with a continuous system of measuring the quality
(reliability) of routine inspection result - in short, to give the
foreman the laboratory control chart as an efficient, low cost "extra
man" to monitor the work-quality of the personnel and equipment turning

1

out these results. The important feature of this philosophy is that, contrary to some of the quality control programs described in recent literature, the laboratory foreman did not become the object of control, but rather became responsible for it. Under these circumstances, the overall "quality consciousness" of all laboratory personnel was improved and maintained.

Foreman Training Course

Since the responsibility of laboratory work-quality rests with the foremen, it was decided to concentrate training efforts on this group. A curriculum was designed to instruct these men in the theory, mechanics and necessary calculations to set up and maintain laboratory control charts; a copy of the program outline may be found in Figure I.

The formal instruction consisted of four 1-1/2 hour sessions. This was handled by two University professors actively engaged in the field of statistical quality control. The lecture presentations were pitched at the practical level, supplying only the minimum of statistical theory. The sessions were spaced two weeks apart to allow practice of the techniques between classes; most of the examples used during class were drawn from current laboratory data. An intensified follow-up program was carried out after completion of the formal instruction; this was assigned to a technical coordinator. It consisted of technical assistance to compile data, set initial control chart limit lines and provide a system of communication to laboratory management on the progress of all active control charts. It is believed that this intensified follow-up was in large part responsible for the continued interest and improved effectiveness of the foremen in translating the subjects of the lectures into practical applications.

Organization and Communication

It is axiomatic that the effectiveness of a quality control program must be maximized by an efficient system of communication from top to bottom of the laboratory organization as well as between the laboratory and its customers. Figure II illustrates the organization which was set up to do this job. Below are described the various activities which have been effective in keeping the program vigorous and continually expanding:

1. Preparation of a monthly summary table on the status of all active control charts by the general foremen of the laboratory control groups to the laboratory head. A copy of this is also sent to the technical coordinator. The assignment to prepare this report has clearly left the responsibility for work-quality in the hands of the foreman and has provided the incentive to keep this vigilance continuous.

2. The technical coordinator calls a monthly meeting of the general foremen, technical supervisors and laboratory head to discuss mutual problems and progress made. Special problems concerning specific tests are brought to the attention of the foremen by the technical coordinator. This meeting has served to tie together the individual programs in each control group as well as to keep management directly informed as to the progress of the overall program.

2

3. Each laboratory control group places the current maximum allowable testing error on the daily summary inspection report sent to the process operating personnel; these statistics are supplied by the technical coordinator. Specifically, each inspection reported to process personnel is accompanied by a statistic defining the maximum allowable difference between two independent tests on the same sample. This difference is the standard deviation multiplied by 2 2. This communication is considered to be very important, since (1) it relates testing reliability to a specific result received by process personnel in non-statistical language and (2) it provides an on-the-spot basis for judging the wisdom of "another sample" if the first result reported is not a "desirable" one. Our records show this action has reduced the number of second and third samples requested by process personnel to "check the first figure".

4. The technical coordinator prepares a complete quarterly tabulation of current standard deviations of test methods covered with control charts. This letter has been effective as a guide to technical personnel in making decisions as to the validity of product quality differences reflected by plant-test data. In addition, it provides reliable work-quality indexes of our laboratories which may be compared with those of other company laboratories.

Procedure of Activating Control Charts

A general, step-wise procedure has been followed in setting a control chart into routine operation. The following describes the sequence of events that has been followed:

1. Assignment to activate a control chart for a specific test method to a foreman by the laboratory head.

2. A system of accumulating data is set up by the foreman. This is designed so that control chart limit lines can be set in about $2 - 4$ weeks.

3. Control chart limit lines are calculated and drawn. The chart is displayed at the site of the equipment used by the laboratory technician. This step usually involves the services of the technical coordinator of all chart activity.

4. Subsequent analysis for control chart purposes are scheduled and the data are plotted as accumulated by the foreman. It is this activity by the foreman that has been responsible for improvement in test reliability. With the chart displayed on site, a simple observation of the chart by the foreman for out-of-control points has made rapid corrective action possible. Since points appear on most charts at least once every two days, changes in work-quality due to different technicians have permitted judicious selection of technicians requiring further training. Enforcement of adherence to standard procedures and use of proper equipment have also resulted in precision improvement.

3

5. Control limit lines are recalculated once a month. Changes in level of precision are relayed to the coordinator and reported once a month to laboratory management.

Figure III illustrates the typical history of a control chart. For purposes of illustration, average and control limit lines have been drawn for all the data (covering about 5 months). Figure III illustrates one of the two types of charts employed - the "standard sample" chart. Both this type and the "range" chart are employed by the laboratories:

(1) The standard sample control chart makes use of two simple graphs-one for a measure of the changes in level of results with time, the $\bar{x}$ plot, and a second for the measure of the spread of results, the R chart. The abcissa for both plots is time, the ordinate a quality characteristic; for refinery products it is normally a specification (sulfur content, flash point, viscosity, etc.). A practical example of this type of chart is illustrated in Figure III. The test method in question had been in use for a number of years and was thought to be very reliable. As the data from blind and random submission of a standard sample began to show a stable level, but undesirable ranges (from 3/29 to 6/4/54), it was concluded that a change in technician was desirable; an examination of equipment and procedure was also instituted. The result of this short-term study by the foreman showed the equipment to be the important factor; in this case a change from a porcelain to a platinum crucible to carry out our ashing procedure was instituted. The out of control points No. 11 for ranges and No. 12 for levels were attributed to poor equipment and an unwise change in technicians respectively. With a change in equipment and the original technician, subsequent ranges (points 15-24) reflected an improvement in generally lower and more stable ranges; the level of results during this time reverted back to that reflected by points 1 through 6 (3/29 to 5/13/54).

The above experience is illustrative of the typical results of employing the control chart technique. Examination of chart information normally reveals that for a period of one to three months a standard well-established test method seems to behave as expected; enough time must be allowed for all possible variables to exert their influence on the overall variability. The test method used in the example continued to reflect ranges higher than were originally thought to be normal before any control chart examination was undertaken. Currently, intensive procedure with a modern instrumental technique.

(2) The second type of chart, illustrated in Figure III-A, is a plot of Ranges (R) only. This is used to monitor methods employed to analyze products which are not stable enough to keep on hand in the form of standard samples. In the example (Figure III-A), a reduction in range (R) with time reflects an improvement in precision. Control limit lines were reset to conform to a lower average range toward the latter part of the time interval (2 months) covered by the chart.

4

Accomplishments

Over a two-year period, the quality control program described has made it possible to place 75 - 80% of the routine inspection work load in statistical control. This includes practically all specification quality evaluations carried out by the laboratories. It involves over 100 test methods, many of which have been improved and all of which have arrived at a state of "statistical control" (consistently reliable at a specified level of precision). It is emphasized that this was accomplished as a result of the efforts of laboratory technicians and foremen as a part of their daily work routine.

Figure IV lists typical improvements that have been experienced with about 90% of the test methods monitored for quality with laboratory control charts. Normally, a reduction to about one-half of the before-chart standard deviation is experienced; some of the tests improved to as much as one-fifth of the before-chart standard deviation.

The program has worked very successfully from the standpoint of calculated monetary returns due to improved test reliability. Perhaps a more important advantage, although intangible, has been a reaffirmed interest of laboratory personnel, foremen as well as technicians, in the significance and quality of their efforts.

Most test methods monitored with control charts arrive at a stable level of precision. This condition represents the best practical level of precision that can be expected to result from routine use of control charts. It also represents a history which serves as a reliable starting point for technical analytical development effort, if the level of precision is still not satisfactory. Some tests show no improvement after a control chart has been in use for several months; it is our experience that such tests require a major change such as an entire new procedure, if improvement is economically justifiable. These experiences clearly indicate that the control chart approach has accomplished two things: (1) it has afforded a very effective and inexpensive method of improving laboratory testing variability for many tests at one time and (2) it has served as a guide in deciding where concentrated analytical development effort should be applied. It is only after these two areas of activity have been exhausted that multi-testing is considered as a means of improving the reliability of results reported by the control laboratory.

Other Benefits

Among the many decisions a laboratory foreman must make every day, it is in his domain to decide whether a sampled product is on or off specification; more generally, to decide what the quality level of the product is on the basis of a single test result. Daily decisions are made by these men whether to (a) report the original result, (b) request another sample or (c) retest the same sample and report the first or second result or average these results. It is in this area that a strong statistical quality control program offers the foremen the "basis for decision" in their daily discussions with process personnel concerning the "true" quality of the product inspected. It offers him valid arguments to counteract coercion by process personnel to (a) retest a sample when actually not warranted, (b) accept a second, third and at times as many as 6 - 8 samples in order to obtain a "desirable" test result.

5

One of the benefits resulting from the use of control charts on
test methods lies in the fact that practically every test so monitored
soon exhibits "statistical control"; this is considered an improvement
even though the data do not exhibit a reduction in standard deviation.
Under these conditions, every single test result is the most probable
candidate for the average product quality measured; in addition, definite
statements can be made as to the maximum allowable difference between
two independent test results on the same sample. It was with these
considerations in mind that a "Decision - Function Policy" for laboratory
foremen was formulated. Figure VI is a copy of the official policy. It
is obvious that without reliable work-quality indexes (standard devia-
tions) of test methods such a policy could not be formulated. It has
been very well received by foremen, and has saved time and money in
making decisions concerning intermediate as well as finished product
quality. The enforcement of this policy has magnified the fact that the
quality control function is a vital part of the laboratory foremen's
responsibility.

Economic Evaluation of Accomplishments

Multi-testing has been suggested as a means of improving routine
testing variability, especially for those test methods which have a high
economic value. The cost of such a program is defined as the number (N)
of replicate tests that must be averaged to reduce the variability
(standard deviation) of test methods to a desired level. The number of
replicates that must be run is calculated as the square of the ratio of
the current standard deviation to the desired standard deviation. This
number (N) is calculated from the fundamental relationship

$$\sigma' = \sigma_{\bar{x}} \sqrt{N}$$

$$N = \left(\frac{\sigma}{\sigma_{\bar{x}}}\right)^2$$

Where σ' = standard deviation of a single determination

$\sigma_{\bar{x}}$ = standard deviation of the average of N
replicate determinations

On this basis, four replicate analyses must be performed and averaged to
be equivalent to a reduction to one-half of the before-control chart
standard deviations; Figure IV illustrates the reduction of the before-
chart standard deviations which resulted from the control chart approach.
It is on this basis that the economic accomplishments of the program has
been calculated. A sample calculation may be found in Figure V. Since
a large number of tests have shown improvement similar to that illustrat-
ed in the example, the incentive to use statistical quality control runs
into the thousands of dollars per day and into the millions of dollars
per year. Of course, it will be clear that a multi-testing program on a
large scale would never have been undertaken. Nevertheless, the magni-
tude of the high cost of multi-testing compared to the use of control

6

charts clearly illustrates the wisdom of going this route; in fact, any alternate route would present a prohibitive cost.

Summary

The quality control program described in this paper has been a successful venture for two years. The laboratory control chart has provided realistic measurement of prevailing testing variabilities for many standard well-established test methods. In many instances, specific and removable causes have been detected and eliminated or at least minimized after isolation and correction. In addition, the quality control program has served to maintain levels of performance and ranges of variation. Most important, the quality control system described has provided a continuous, representative check on overall laboratory performance, and therefore reflects management acceptance of its responsibility for maintaining standard and stable conditions. As a result, the huge inspection job of our refinery today is being run like an efficient business. The commodity we sell is data. Our aim is more reliable data for the customers we serve. Quality control has been an effective means for doing this job.

FIGURE I

TRAINING PROGRAM OUTLINE

First Session: Fundamental Concepts

    a. demonstration with models and gadgets
    b. definition of "quality" and"statistics"
       implied in the title
    c. patterns of "normal" variation

Second Session: Types of Variation

    a. frequency distribution
    b. demonstration of kind of data used
    c. concept of "population" and sampling for an
       estimate of variation
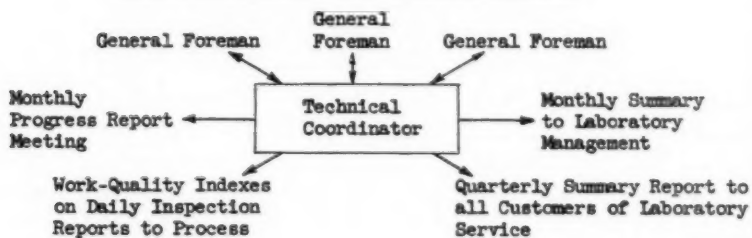
Third Session: The Control Chart for Variables

    a. application in the laboratory
       1. standard sample
       2. for duplicate analyses (one example for
          each type)
    b. applications in the plant
       1. demonstration with case history

Fourth Session: The Control Chart (continued)

    a. for moving average and range
    b. discussion of mechanics for maintaining charts -
       illustrated with advanced case histories

FIGURE II

ORGANIZATION FOR EFFECTIVE QUALITY CONTROL

General Foreman   General Foreman   General Foreman

Monthly
Progress Report
Meeting

Technical
Coordinator

Monthly Summary
to Laboratory
Management

Work-Quality Indexes
on Daily Inspection
Reports to Process

Quarterly Summary Report to
all Customers of Laboratory
Service

9

FIGURE 3

10

FIGURE 3A

11

FIGURE IV

TYPICAL REDUCTION OF STANDARD DEVIATION
AFTER USE OF CONTROL CHARTS

Test No.	Std. Dev. Reduced by a Factor of
1	0.50
2	0.25
3	0.25
4	0.14
5	0.16
6	0.33
7	0.50
8	0.33
9	0.25
10	0.50
11	0.25

FIGURE V

SAMPLE CALCULATION OF DOLLAR SAVINGS
DUE TO REDUCTION OF STANDARD DEVIATIONS

Test Method:	Vapor Pressure
Unit Cost:	X Dollars *
Avg. No. tests run per day:	13
Std. Dev. before control chart:	1.0
Std. Dev. after control chart:	0.33

$/day

Present cost to handle workload:	13X	
Cost of control chart analyses:	X	
Total cost of present system:		14X
Cost of multi-testing: (9) (13) (X)		117X
Dollars/day saved:		103X

*Unit costs vary from 2 to 20 dollars

13

FIGURE VI

DECISION - POLICY

FOR

REPORTED LABORATORY INSPECTION

Introduction

The procedure outlined below is designed to serve as a guide to laboratory foremen in making decisions concerning the question of whether or not to report a test result on a sample. Its purpose is to enable foremen to make this decision on a maximum of two samples.

Procedure

Case I: When current control chart data shows test method employed to be in statistical control.

    1. Report the result as is.

Case II: When current control chart data shows test method employed to be out of control.

    1. Re-test independently on same sample; if the difference between these two results is within maximum allowable error, report the average of the results; if the difference is greater than the maximum allowable error, discontinue analysing routine samples until the method is known to be in control.

    2. If not enough of original sample is available for second analysis, request a second sample of sufficient size to make two determinations. If the difference between the first analysis of the second sample and the original result is within the maximum allowable error, report the average of these two results.

        If the difference between original result and first analysis of second sample is greater than the maximum allowable error, analyze the second sample a second time. If the difference between the two results on second sample is within the maximum allowable error, report the average of these results; if the difference is greater than the maximum allowable error, discontinue analyzing routine samples until the method is known to be in control.

Case III: When there is no control chart in service for the method employed.

    1. When the test method employed is believed to be unreliable, obtain two independent test results on each sample. Report the average of these results.

When the range (differences) between at least 15 duplicate tests, exhibits statistical control, analyze samples only once and report that result.

MANAGING THE QUALITY CONTROL FUNCTION

Robert G. Mitchell

International Latex Corporation

INTRODUCTION

In order to discuss a subject as broad as the managing of anything, certain restraints must be drawn around it so that the discussion may have some fruitful meaning to the participants. I would, therefore, prefer to generally express some thoughts on managing the function in which all of us are apart rather than trying to give anyone the opinion that here is a complete rundown on the entire topic. Moreover, I will confine these remarks to our overall company situation rather than use any outside criteria.

It is my opinion that before we can proceed with anything else, we must understand the company atmosphere around us and our place in it. As a staff organization within Manufacturing, we are responsible for serving a line group on all quality matters. We are dedicated to assisting line management through close teamwork to successfully manufacture a product economically that conforms to raw material and product specifications drawn up by Research and accepted by the line in Manufacturing.

If we agree that this is the situation in which we find ourselves in broad terms, then it is the "how" of assisting line management effectively that we really want most to direct our attention for this discussion on "managing".

Each of us is a manager of all or part of the Quality Control function in his particular plant. Each of us has problems in common regardless of plant location or product differences. The art of management, while it must fit the conditions of the plant and products concerned, still remains a basic field unto itself. We can manage well, we can manage poorly, or we can just struggle along; but to really succeed in our work in Quality Control, we have only the obvious alternative!

DEFINING THE FUNCTION

Each of us has been given and understands the objectives of Quality Control in his company. We all operate under a similar scope of function. These objectives and scope of activity provide us basic guides for operating. They are prerequisites for managing, for without them we would have no purpose or road to follow.

Beyond these bare minimum needs of direction, we must develop a means for fully conducting those activities for which we are responsible. This must come from "planning" the role to be played and not to be played by Quality Control in its operation. Such statements do much to clarify with others as well as ourselves the Quality Control job more specifically than the broad terms expressed in the scope of the function.

Basic concepts must be established to be used as a kind of motivating theme. Among these concepts would be these:

1. Use facts, not opinions
2. Let the facts supply conviction

17

3. Concentrate on management controllable defects rather than operator controllable defects
4. Work on the major few defects rather than the many minor defects.
5. Dollars saved are generally more important than percent reductions
6. Maintain a balance of effort between chronic and sporadic defect studies.

USE OF MANAGEMENT TOOLS

All these items mentioned thus far - objectives, scope of function, role to be played, basic concepts - are identifiable as guideposts that mark the way for us. They provide us with the rules of the game. But now we must play according to them to achieve results and we must look for tools now to make job success a daily feature instead of a grand occasion.

The need for a solid, tight organization is one of the fundamental of such tools. Whether it involves a single engineer or a group of engineers, the problem of establishing people in clearly defined positions effectively is the same. The organization once established requires review often for it must be altered to fit constantly changing conditions. Flexible, well defined organization of the function is a must.

Administrative tools are also fundamental for they steer each manager along the way. The creation of annual plans, quarterly reviews of past performance and future plans, expense budgets, cost reports and other aids like them provide guidance for an ambitious, balanced program of Quality Control activity. Lack of balance is one of the greatest hindrances to overall accomplishment there is!

The creation of procedures and methods of handling the specifics of the function are most necessary for efficient operation. I believe in the philosophy that it it hasn't been written, it hasn't been well thought out. While this may not always be possible for all things, it is a fundamental need for most. Misunderstanding instructions is a common occurrence, yet could be greatly minimized by a concisely well written confirmation. Misunderstandings delay results and confuse; procedures expedite results and gain uniformity of action and interpretation.

With our function as new to our company as it is, we have much training of all people to do so that we may satisfactorily do the Quality Control job. Verbal training alone and disjointed will never yield entirely satisfactory performance. Training programs for key work must be developed so that instruction is uniform and requires proof of reasonable understanding by the recipient.

We use statistical tools in our work. These are the tools that tend to mystify all but ourselves - and even we experience distress with some of them. Use of statistics must be made wherever they will substantially assist in getting the true facts economically and enable us to offer better recommendations to the line. We should give these tools full expression yet be sure we speak to others in the laymen's language so that their benefits will not be stifled by the lack of technical understanding on the part of line management.

The development of goals requires also the development of yardsticks and scoreboards with which to evaluate achievement. Whether these be quality performance, vendor conditions, dollar savings or any other form of study, we must create means of measuring performance. How else can work effort and accomplishments be reliably identified? Providing these adequately will give added guidance to both the function itself and to line management as well.

Above all these tools, we must keep in mind that uniform raw materials, reduced defectives and better quality going to the customer must result in dollar savings at the same time. No proposal will be well received by the line unless it has a cost consideration behind it. Dedication to saving dollars while improving quality conditions is our only acceptable and correct approach to serving our plants well.

THE SPIRIT OF MANAGING

So far, I have confined my remarks to those basic requirements and tools that can be and have been said over and over again. To me they are the surface story of managing. Without any one of them, the function will not realize its full potential. I can recommend them as vital for the capable managing of our function.

But the key to real success in managing as I see it, lies in a more personal vein. I see the key in the "spirit" of the individuals participating in Quality Control work from the manager spreading throughout the organization. I say that because successful managing requires so much more than just the rules as they are so often expressed.

Each of us must provide personal leadership and inspire confidence in all people with whom we are associated. Our enthusiasm must be catching. The spirit of achieving hard won success as a team together with other line and staff people must prevail. There is no place in any of our plants for a pole to carry the Quality Control flag higher than those of the other plant groups.

We must sell our services to the line and convince them that they need our brand of service. If we are to be used correctly and effectively, we cannot demand but must prove by our achievements that we are the best at our kind of work.

We must be self starters! We don't wait for assignments, but dig in and do something about every situation as it comes up that is in our line. And we don't ignore it if it just doesn't fit us to a "T". We discuss it with others and only bow out if some other group is better equipped to handle it. In other words, we don't have and cannot have a wall around us that keeps us confined with blinders. Each plant has a team; Quality Control must be an active member!

We serve the plant manager well by assisting Production in every way possible. The planning of a balanced program of activity is always a joint effort with Production and other staff groups that are concerned. Revisions in basic plans likewise are always done in consultation with the other groups so that smooth functioning of the team unit remains through discussion and communication.

Good managers are not "fence sitters" either. No one expects rash

19

decisions made without due consideration and fact finding. But the managing of any function requires a positive, confident approach. We cannot serve the line adequately by continued reservations to issues. The line takes the risk in making decisions and they depend often on the recommendations of staff. Consequently, it is our job to suggest at least one proposal and preferrably several alternatives where possible to the line by the deadline time for a given decision to be reached. In carrying out our functions, we should always submit a recommendation along with any problem that arises. The plant manager, Production and other service groups must know where we stand!

It is appropriate that we should also give attention to the problem that arises seldomly, we hope, but none the less does happen. That is the occasion when the line chooses to take action contrary to our suggestions. Under these conditions, a good manager really shows up!

Surely we must believe our proposals to be correct when we give them. Moreover, we must continue presenting our point of view in a variety of expressions of facts until we have gained our point when we know our proposal to be right. But let us be certain all the while that we listen to and respect the other points of view; that we fully appreciate the prerogatives of the line to make the decisions for which they are responsible; and, above all, support the decisions finally made with all of our capabilities! Such action on the staff manager's part is one of the most difficult to learn and one of the most necessary personal ingredients to have for consideration as a good manager.

Just one other thought on this situation. Let us always respect the line management and their positions of pressure. If we respect ourselves and our profession, we must likewise show in full return a genuine respect for the line. When things don't go just the way we have recommended, it is far better to first look at ourselves and what we have proposed again before casting an eye in the other direction!

THE OTHER SIDE TO MANAGEMENT

With all of these thoughts on managing that I have expressed in the form of basic concepts, tools at our disposal, general basic rules of operation and the "spirit" so necessary for effective achievement, there is one last but far from least consideration that I would like to leave with you.

Ours is a job of technical staff work. Quality Control is a new function operating in new plants with new men in our company with line men who, for the most part, are unfamiliar with the techniques of the function. These line men are eagerly looking to us to be of real help to them and we likewise want to be of help as quickly as we can. Such an atmosphere has afforded us every opportunity to put into practice many systems and practices so far with great potential for success.

But much is at stake. The line is looking to us for competent, successful activities both technically and generally. They look to us to manage well in all things.

This entire discussion has primarily been slanted to the technical side for a purpose, yet there is another side often taken for granted that never can be ignored as "in the bag". This other side consists of such things as reporting to work on time, a full 8 hours work for each

person in the group, breaks that do not exceed the plant policy, reports distributed on time, good human relations practices, etc.

There are many items to be included, but I should like to stress one in particular. That one is the art of supervision. A plan or system can be installed in the plant that can save money and provide better control than ever before, yet fail because it wasn't supervised well. All of us can readily agree that this can be true, but so often we shrug it off as insignificant because "we don't let that happen to us". Other people have that problem, not us!

My word here is one of caution. We are a new function. We know it is justified as a part of the manufacturing team. We know what great potential good can come from its practice.

Let us not then ever feel complacent about this so-called "other side". Much of the good in the technical side will go astray in the eyes of the line and the people in the plant if these other side necessities are neglected. Almost any system well taught, well supervised and followed up is to be preferred to a better system that is not!

SUMMARY

We have covered generally a broad view of some management needs that have application in many fields other than Quality Control. Defining the function well and using various management tools properly applies to all groups. What I have tabbed the "spirit" of managing is perhaps even a way of life if looked at in one sense. There is no question about the need for the "other side" of management to stand as the basis for all technical activity.

I have tried to stress that we must be especially mindful of conducting Quality Control soundly because we are so new to the manufacturing scene. In order to remain, we must prove that our presence materially contributes to a plant's progress toward optimum efficiency and low cost. We must be "good" to achieve it!

A good, balanced Quality Control Program is not an accident! It just doesn't "happen"! It must be planned! It must be well managed!!

STATISTICAL METHODS IN ENGINEERING DESIGN

Enoch B. Ferrell
Bell Telephone Laboratories, Inc.

ABSTRACT

Variability in a manufactured product will probably always be with us. We can live with variability if it is predictable, i.e., if it is self-consistent and stable. To achieve stable variability we need a proper coordination between the design of the product, the design of the process used to produce it, and the components and tools used in the process. The simple but effective techniques of statistical quality control are powerful aids in arriving at this coordination.

- * - * - * - * - * - * - * - * - * -

When we start to talk about the business of design, we should realize that it has several objectives. We should design a product that will do what we want it to do after it is made. We should design a product that can be made at a reasonable cost. We should design a product that will perform its desired function not only when new but also after considerable use or life. And we should design a product so that if something does go wrong with it we can get at the trouble to fix it. This last, which we might call "Design for Maintenance", in some cases reduces to the simple ability to remove one unit of an assembly and replace it with a new unit.

If we study these various objectives, we soon discover that reproducibility underlies all of them. When we say we want to design a vacuum tube for a certain use, we really are not talking about "a" vacuum tube at all; we are talking about many vacuum tubes to be made in the future according to this design and we want all of them capable of performing the desired function. That is, we want interchangeability, which to the manufacturer means reproducibility. When we talk about "life", we mean not the life of a few tubes tested in the laboratory, but the life of many tubes which will be tested in service. Here again we are interested in the reproducibility of that life. And when we talk about maintenance, which very often involves replacement of piece-parts or sub-assemblies, we want uniformity in these replacement parts.

If we accept the thesis that reproducibility is an underlying objective of design, we find it very difficult to separate the design of the product from the design of the process that is to produce it.

We are not likely to get very far in the design of a new product and its associated process without wanting to turn to the laboratory or to the shop for more information. We want to know the strength or thermal expansion or some other property of a proposed material. We want to know the tolerance capability of some tool. So we plan experiments and make measurements. And when we start to interpret the resulting experimental data we immediately raise the questions of credibility, reproducibility, precision, and the like as applied to the experiment. Some of the best tools to couple with our engineering know-how, and to help us answer some of these questions are the simple, but effective, statistical methods that have been developed by the Quality Control Engineer.

We sometimes hear that the methods of statistical quality control are applicable only after we are in large-scale production. Many of us, however, have witnessed the great assistance that these statistical techniques can provide in the initial setup and adjustment of our mass production processes. It is generally recognized that these techniques are of enormous value in the pilot shop. It is not generally recognized just how early in the history of a new product the application of statistical quality control methods should begin. Let me cite a few examples.

This first example goes back about to the beginning of my contact with statistical methods. At the time we were working on a glass-sealed switch. Our work resulted in the design of a product and a process for making it. This glass-sealed switch is a very simple device. It consists essentially of two plain permalloy reeds supported at their outer ends with their free ends projecting towards each other cantilever fashion with a little overlap. See Figure 1. When placed in a magnetic
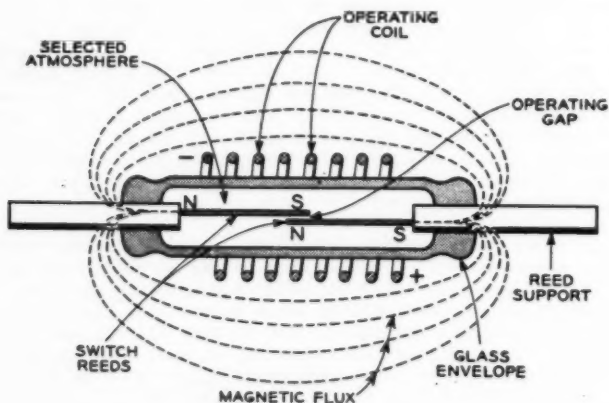
Figure 1 - Glass-Sealed Switch

field the free ends attract each other, come together, and make contact. The supporting structure is a glass tube surrounding the reeds, and sealed to the reeds at their fixed ends so as to enclose and protect them. One of our early problems was the size of the magnetic gap between the free ends of the reeds. How much should the reeds overlap and what should be the spacing through which they moved to make contact? We set out to determine an optimum experimentally, but found it very difficult either to produce or to measure a given set of dimensions. We also found our results influenced by the length of the reeds, the type of seal which held them and several other things which seemed to have a great deal of variability. It appeared that we would have to design the process before we could complete the design of the product.

At about this point we arranged for another department of the Laboratories, acting more or less as a sub-contractor, to produce a small number of these switches on an experimental basis for our study. Their people prepared a set of tools, set up a process and made some switches.

24

After some experimentation on their part, they announced that they were
ready to make experimental models. We gave them a set of dimensions and
they made 48 switches.

We received these 48 switches, measured certain characteristics,
including the ampere turns required to close the contact, and plotted a
statistical quality control chart of these measurements. See Figure 2.

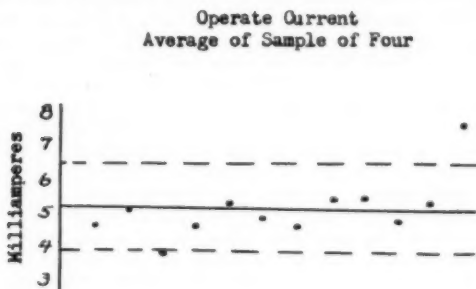Operate Current
Average of Sample of Four

Figure 2 - Operate Current. A study of 48 switches
made under supposedly constant conditions,
using the "Old Tool".

Each point on the chart is the average operate current for four switches.
The limits — the dotted lines — are computed by a standardised sta-
tistical method. The location of the last point outside the limits on
our control chart indicated that the last 4 switches were fundamentally
different from the first 44. We went to our sub-contractor and inquired
about this, but were assured that all 48 switches were just alike. We
protested that our computations indicated a fundamental difference and
he asked us to check our computations.

We did check our computations, found them correct and set about a
two weeks program of study in an effort to explain the outage. At the
end of this time we had discovered that whereas the magnetic reeds for
the first 44 switches had been cleaned by heat treatment in a hydrogen
furnace, the reeds for the last 4 switches had been cleaned in an acid
bath. This had a significant effect on the surface condition and hence
on the stiffness of the reeds. We had detected the presence of an
assignable physical cause for abnormal statistical variation and had
actually identified the cause.

We showed the sub-contractor our findings and the difference was
readily admitted. Here is a very common pitfall. A change had been
made in the honest belief that it would have no effect at all. But it
did have an effect and the simple statistical analysis of actual results
had shown that it did.

We were now no longer interested in the last 4 switches. We re-
plotted our control chart for the first 44 switches and found all points
now inside the control limits. We sometimes describe this situation by
the phrase "in control". The indicated tolerance, however, was some-
thing like ±25 ampere turns. As compared to an average of about 50,
this was entirely too much variation for our purposes. Our

25

sub-contractor pointed out that although the variability was large, he thought he knew why. One of the supports in the assembly tool was a little weak and could be stiffened. Another part of the assembly tool obstructed the vision of the operator; and this could be corrected. "This should reduce our variability".

Our statistical analysis, however, had given us another piece of information. The 44 switches appeared to be in control. Now when we have a state of control the variability is not due to one or two important causes, but to a great many unrelated minor causes, and the removal of one or two of them will not make any important reduction in the variability. This, in fact, is what the words "in control" mean. We recommended that he not spend time on these minor modifications in the assembly process. We said that in order to significantly reduce the variability this whole process must be heaved in the ashcan and a completely new process with a new set of causes for variation must be designed. Our sub-contractor, however, still thought it would be worthwhile to try to fix up the existing process. It is common at this point to have some discussion as to possible meanings of such phrases as "minor modification of the process", "major redesign of the process", and "brand new process". These semantic difficulties are seldom resolved at this time.

During the next six months, the sub-contractor made several modifications of the assembly tool and produced a considerable number of switches. A record of this in the form of a control chart is shown in Figure 3. The part at the left is a history of what he did before our discussion of the 44 switches. The last 11 points in this section represent the 44 switches. We have labeled this section of the chart "Old Tool". The next section of the chart which we have labeled "New Tool" represents his effort to modify the process so as to reduce the variability.

Briefly, the upper chart is a control chart of averages. On it each point represents the average operating current for each successive sample of 4 switches. The line through the center is the grand average and the dotted lines are control limits. They are computed by well known, simple statistical techniques and represent the limits beyond which we do not expect a respectable process to go. You will note that there are many points outside the limits. Actually, these points represent, for the most part, deliberate readjustments by the operator. He would make a small lot of switches; he would observe that the average operate current for this bunch of switches seemed to be low; and he would adjust the tool upwards. More than likely the next bunch of switches would possess a high average operate current and the tool would be readjusted down again. Hence, on this chart the control limits represent the inherent capability of the process and the outages represent the human interference. The lower chart shows the variability or dispersion. Each point represents the variability or spread within a group of four switches. It has no points out of control. This means that the process in use had a stable inherent variability. The operator's readjustment had little or no effect on this. The process was inherently stable and this was its dispersion. The middle section of the chart indicates that by means of the new tool the variability had been reduced by some 15%. This is not much of a reduction and actually the total amount of data available here is not enough to be sure that this improvement is real.
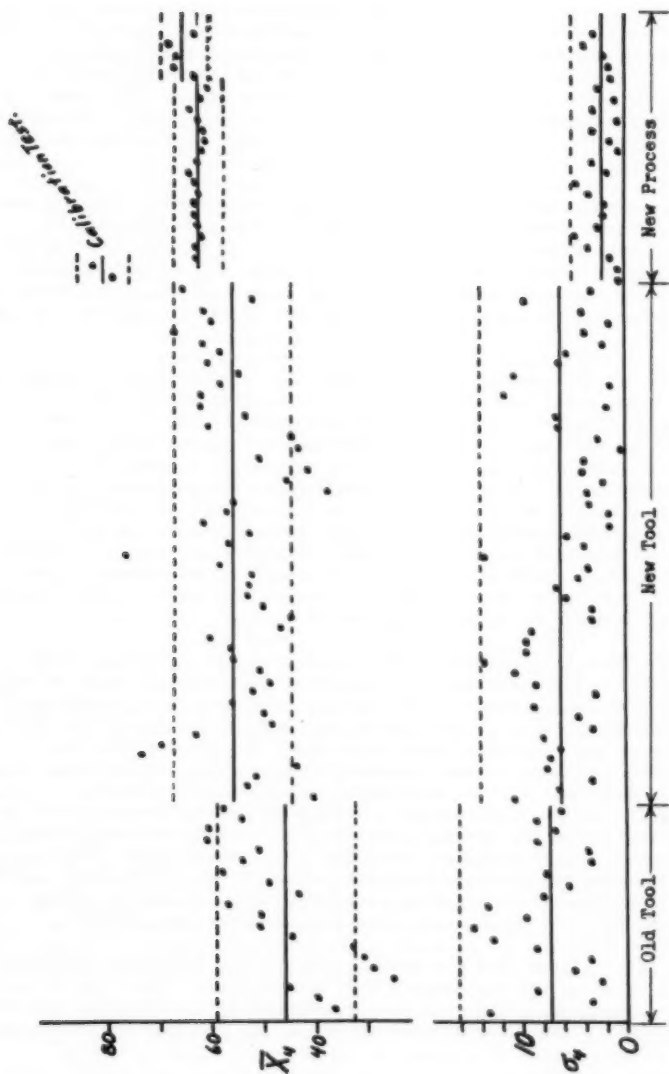
26

Figure 3 - Operate Current. A comparison of minor modification in the process with introduction of a new process.

27

During the time that the sub-contractor had been working with the new tool, one of our engineers had been groping for a new assembly process. He proposed several different processes. Some reason for dissatisfaction with each of these was discovered before it was put into use. Finally, a new assembly process was conceived and tools for it were built. The two points on the chart at the beginning of the section marked "New Process" — the two points labeled "Calibration Test" — represent the first 8 switches made by this new process. With some optimism the indicated low dispersion was accepted as valid. The average of the 8 relays was accepted as indicating the actual tool setting, and a single correction on the tool setting was made.

Switches represented by the rest of this "New Process" section were then made with the results shown. The control was good and the dispersion had been reduced by a sizeable factor. The last little section at the end of the chart shows the effect of using a different magnetic material for the reeds with no other change in the switch or the process. The dispersion appears to be not affected; the average is affected only slightly.

Let me review some of the things that this case history illustrates. First, it illustrates the ability to find assignable causes for variation — even though, in this case, the assignable cause arose from a process change believed to be so unimportant that it was overlooked. Second, it illustrates the fact that if a process is essentially in control, tinkering with the process, modifying it here, polishing it up a little there, will have no real effect on the inherent variability. When we combine these two, we realize that before you can evaluate the design of a product or a process you must bring the process into control. If, after it is in control, the process displays an intolerably large variability, you might as well heave it out and start over. You will not improve the variability by modifying the process.

We have been talking here about variability and stability in the same breath. It may seem that the two are incompatible but this is not so. The instantaneous value of the voltage delivered by an oscillator has great variability — we say the voltage varies sinusoidally — but the oscillator may have very great stability. This means that the magnitude and frequency of the variation is constant. The concepts of variability and stability can be applied in combination also to resistance noise. The instantaneous voltage of the resistance noise is highly variable, but the average power at a given temperature, bandwidth and resistance is quite stable. When the statistician talks about statistical stability or when the quality control man talks about a state of control, they are talking about a variability which may be either large or small, but a variability that is consistent with itself.

As a third lesson from the example look at the quickness with which the capabilities of the new process were established. There was no lost motion in demonstrating that the new process was an improvement. We had a good measure of the inherent variability of the old process to compare it to.

I might mention that when our sub-contractor saw the new process he quickly agreed that we were not stretching the meaning of the word when we said the process was new. It bore no similarity at all to the process whose variability we had found undesirable earlier. I might also add

28

that since that time — and, I hope, possibly to some extent because of that experience — that sub-contractor has become a staunch supporter of the statistical quality control idea.

Here is a second case history. We were working on some similar glass-sealed switches which contained little pools of mercury that wetted the contacts so that the electrical circuit was closed and opened by liquid mercury. At one stage during our work on these switches a lot of some 30 or 40 were made and their operating characteristics were measured. This was finished up late in the afternoon and the switches were put on life test to run overnight. I took the data from the measurements home with me that night and on my knee during a 35-minute ride in a commuters' train I analyzed these measurements by use of a simple control chart. The next morning, almost before I got to my desk, the engineer who had made these switches and placed them on life test came to me with the statement that one of the switches had failed. It would not even run overnight. I asked him which one. He told me. I checked with my control chart, and found that it was one of those whose measurements had been out of control on the previous afternoon and gave him the identification of four or five other switches which were also out of control and might be expected to misbehave. All but one of these other switches failed before the day was out. Most of the remaining switches gave lives that were measured in weeks.

This second example illustrates the fact that when a control chart says an item is fundamentally different from the bulk of its fellows, this difference is worth some attention.

For my third case history let me move on in time to where we were setting up a pilot shop to make these glass-sealed switches at a rate of a few thousand a month. This rate of production did not warrant any fancy tooling-up job, but did justify making a set of simple, easily operated non-automatic tools. The design of the switch was such that a large part of it was fabricated by welding together small pieces which had been sheared off from wire or strip stock. We made a set of simple hand operated shears to cut off these pieces. See Figure 4. We could change from one piece-part to another simply by changing the die and stop in the shear. The shears looked simple, reliable, and easy to operate. As they were set up we sheared off 24 pieces of each type of piece-part, kept them in order, measured them, and plotted control charts. We also measured a random sample of 4 from each of 4 succeeding lots of 80 pieces. We found that all but one of these shearing operations indicated a close approach to control and indicated a uniform spread of a little less than ± 1 mil. The similarity of the variations of the different types of piece-parts gave considerable confidence in the stability of the shear and its suitability for a variety of jobs. Since we had determined its inherent variability by a study of nearly 500 pieces — 40 from each of a dozen types — we had considerable confidence in this figure of ± 1 mil.

The departure of the average, in the case of each type of piece-part, from the intended dimension indicated that our set up error was also roughly ± 1 mil. Accordingly, we marked these piece-parts on our design with tolerances of ± 2 mils and prepared inspection gauges to suit.

The one piece-part which did not show the close approach to control was a piece-part not really suitable for this type of shear. The ability
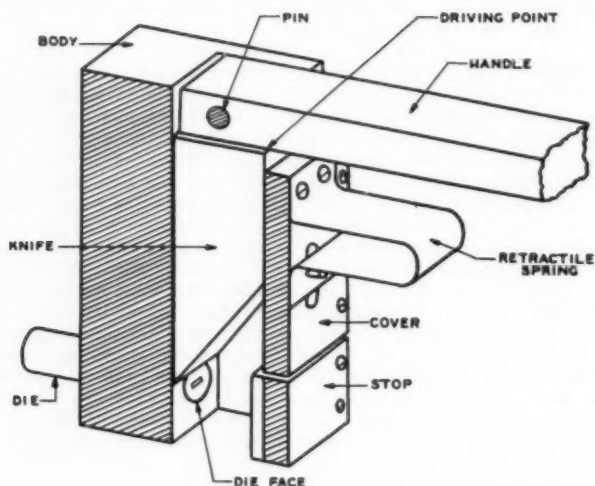
29

Figure 4 - Shear

of the control chart to aid in determining the suitability of a tool for
a job and in determining its inherent variability when used on a variety
of suitable jobs is one of its great advantages.

As a matter of expediency, in this case, we went ahead and used the
shear on the piece-part for which it was not really suitable. But we
did this with our eyes open and gave special attention to the selection
and training of the operator who used it.

This illustrates, I think, how a little bit of experimentation and
a little bit of statistical analysis can supply us with tolerance figures
for our design that are realistic and that fit the design and the process
to the equipment available.

As a final example, let·me go a little more modern. Quite recently
we were building an experimental switching system such as might be used
in a telephone office. The circuitry in this system involved some 2,000
germanium diodes. These were obtained from several different suppliers
in lots of from 200 to 800. As these lots came in, the diodes were
measured. We, of course, had nothing to do with the manufacturing
process and so our measurements had no effect on the manufacturing
process. The purpose of our measurements was not even primarily an
acceptance inspection, although in the end we used these measurements to
do some of this. Our principal purpose in these measurements was to find
out before we completed the design of our circuitry what variability we
should expect in these diodes. We found that the distribution of back
resistance, for example, was a badly skewed distribution which in the
statistical language could be described as something like a Log-Normal
distribution. The variability was so great that it did not make sense
to talk about a "mean plus or minus" something. The limits could be

better described as half a megohm times or divided by 9. That is, we found diodes running all the way from 55,000 ohms to 4.5 megohms. With this knowledge we designed our circuits to work with diodes of 50,000 ohms back resistance. In two years of almost continuous operation we have had practically no trouble with diode circuits.

This last case bears on the problem of the circuit man or his mechanical equivalent, the assembler. He admittedly has little or no control over the piece-parts that he is to assemble. This makes it all the more important that he should make a realistic evaluation of the variability that does exist in his piece-parts and should design his assembly so as to accept that variability.

We have been talking about variability and stability. These are not new concepts. The experienced engineer nearly always takes them into consideration. He may not employ a formal language in discussing them. His own use of them may even be more or less subconscious. But, in the end, he usually does a pretty good job of insuring a satisfactory variability with reasonable stability.

Two points, however, we would like to emphasize: First, these ideas can be formalized. There is a definite and acceptable language to use in discussing them. There are simple computational methods that can greatly improve and sharpen up our engineering judgment. Second, we can profitably use these ideas and analytical techniques in the earliest stages of development, even in research, so that when we arrive at a design for production we have a solid knowledge of what variability we can expect and a confidence in its stability.

31

APPLICATIONS OF STATISTICAL METHODS IN BUSINESS ADMINISTRATION

Dr. Grant I. Butterbaugh
University of Washington

The subject matter contemplated by the title of this paper is nothing if not voluminous. It would take a series of sessions such as this one in order to make more than a slight introduction to its full discussion. Without doubt the applications of statistical methods in business administration are legion. We have ample evidence of that in the writings of the past, but we can only speculate on the many unrecorded uses of statistical methods by the non-vocal or non-literary practitioners. In this paper, therefore, I shall speak of the beginning applications of statistics in business administration, the changing meaning of the term "statistical methods", how scientific development has progressed, what the situation appears to be at present, and the direction in which research is pointed. Furthermore, being in the field of education, I feel it would be appropriate to make a few remarks on the basic educational problem in business administration. I shall not go into the subject of the design of industrial experiments - a vast field with a bibliography of thousands of references; for a symposium on this subject see reference (1).

As you know, the word "statistics" is used in different ways in everyday language and so the applications of statistical methods cover a wide field of endeavor known as business administration. One use of the word "statistics" is in reference to a collection or presentation of information in the form of numerical data. Another use is in reference to a method or technique for dealing with numerical data. In the two aforementioned senses Statistics is a very old subject. Statistical data have been found in the early writings of the Babylonians and Egyptians of some 4,000 years ago when business transactions and accounting records were recorded on clay tablets and papyrus. Census statistics formed the basis for the administration of early states in their collection of taxes, in the drafting and provisioning of armies, and in the business of building vast empires. Many an unearthed temple corner-stone or collection of clay documents has yielded mute evidence of the importance of inventories of men, munitions, and food in the early administration of the church-state. One clay tablet in the British Museum originating about 1600 B.C. is inscribed with forecasts concerning the chances of a sick man's death or recovery; another with the probabilities of the success or defeat of a military expedition. These events were evidently of such a repetitive nature that it was possible to develop some form of statistical prognostication regarding them. At any rate, the history of record-keeping is the history of accounting and statistics, and all of this record-keeping lies in the realm of descriptive statistics.

While the great preponderance of statistical work of the world has fallen, and still falls, into the category of descriptive statistics, the heart of modern scientific statistics is based on the theory of probability. Here we encounter so-called inductive or inferential statistics, wherein generalizations are made from samples of data. Statistical inference based on samples involves the application of probability considerations to the sample data. This is what is commonly meant when we speak of "statistical methods".

Statistical methods based on the theory of probability have been

developed mainly by statistically-minded mathematicians or mathematically-minded statisticians during the past two centuries. These methods have been put to use in all true sciences and in the many arts by applied statisticians and statistical administrators. The application of inductive statistical methods in business administration is of comparatively recent origin compared with the application of descriptive statistical methods. Business Administration appears to be one of the last of the main areas to accept intensive application of probability statistics or scientific sampling. One can only wonder at the possible reasons for this phenomenon, since business administration has been in evidence since the dawn of history.

Perhaps business administrators saw at first only the need to collect and record data as an aid to memory. Later on, the budget was born out of a desire to forecast operations and to set up standards of operation. The development of budgeting brought on advanced accounting analysis, which went from simple comparisons of absolute amounts of change, to the use of ratios, percentages, and index numbers to express relative comparisons. Here analysis seems to have stalled. Although many trained accountants have worked out deep-cutting analyses of variations in costs, revenues, profits, and financial position elements, the theory of probability has not come into general use to explain to management the risks they take in making decisions involving such variations. Probability has not been a basic study in the education for business. Although businessmen take risks in all phases of their business dealings and in making business decisions, they ordinarily have not had the opportunity to learn how to use statistical methods based on the theory of probability. This is largely the fault of education, of which I shall speak later.

Variation is a fundamental characteristic of science, business administration, and of life itself. Variation may be obvious and inescapable in some cases, and readily explained. In other instances it may be quite elusive, subtle, and not easy to clarify. We know that the measurement of any real thing can never be had with exactitude. There can be no human eye that can detect infinitesimal separations. Variation between sample measurements is therefore inevitable and sampling variation or error must be recognized as real. We recognize two types of variation in repeated measurements in business administration, ascribing them to two different sets of causes:

1. **Assignable variation** - brought on by human, mechanical, or material imperfections which may be sought out and removed if it is economically feasible to do so. The improved measurement may not be worth the cost of the effort, but that is a matter of business judgment and decision. Assignable variation is akin to bias and bias results in inaccuracy - missing the target or true universe value. To the extent that a sample result differs from the universe value, to that extent a sample result is inaccurate. One may set up a budget of revenues and expenses, of operating results, or of financial position. Against this budget one can measure variation in the sample results. Again, standards of work output may be set for men, machines, or whole departments or plants against which comparison may be had through sample results. Accuracy can only be measured when a nominal value, a standard, a budget, or some other target is arranged.

2. **Random or inherent variation** - comes about from the infinite

34

causes of variation in materials, methods, men, and machines. Random variation is capable of reduction but not of complete removal. This random variation or error can be measured and controlled but cannot be eliminated completely as long as sampling is used as a method for estimating the universe values.

As one or both types of these variations can be expected in all sampling operations in business administration or anywhere else, it is the function of statistical sampling methods to eliminate bias, measure variation or sampling error, analyze such variation, and in respect to that analysis to provide sound bases for making certain types of statistical inferences regarding the true measures of the target universe. Business administrators should distinguish between the "target" universe value sought and the "sampling distribution" or "frame" value sought. Since we cannot ordinarily measure accuracy we can only do our best to eliminate the multitudinous types of bias in our statistical methods of collecting and analysing data. We can, however, measure precision or sampling error in the sampled estimate of the value sought. This precision measurement gives an indication of the amount by which the estimate made from the sample deviates from the result that would have been obtained through infinite sampling, or complete coverage using the same methods, equipment, time span, and personnel. One can then express his degree of confidence in the sample statistics as estimates of the frame or sampling distribution value. Since business administrators make decisions based on evidence, it would seem advisable wherever possible to have an objective expression of the degree of confidence in the evidence submitted. Scientific sampling methods give answers to some of the businessman's quandaries.

The theory of probability underlies true scientific methods and techniques used by management in making decisions. Research has brought forth a vast reservoir of ideas for business use. This can be illustrated by the development of non-parametric and distribution-free methods of estimation. The idea of ridding oneself of the assumption of normality or knowledge of the shape of the basic universe distribution in sampling has an appeal. Getting to know the vast array of techniques and the usefulness and limitations of these mathematical methods is a broad challenge to the capability and perseverence of any statistician. One article alone had a reference list of some 1700 articles on non-parametric statistics; this is a veritable library of ideas worthy of note. (2)

Over the past sixty years one management movement after another has evolved dealing with the problems of industrial production and business administration. In 1911 Frederick W. Taylor published his book "The Principles of Scientific Management" (3) bringing forth principles in connection with controlled experiments made on metal-cutting techniques at Bethlehem Steel. Taylor summarized his concept of scientific management briefly as:

1) The duty of management is to "gather together all of the traditional knowledge which in the past has been possessed by the workmen and then of classifying, tabulating, and reducing this knowledge to rates, laws, and formulae which are immensely helpful to the workmen in doing their daily work".

2) Another duty of management is the scientific training of workers.

35

3) Management has the duty of co-operation with labor to make sure that the work is done in accordance with the principles of scientific management.

4) It is the duty of management to make an almost equal division of work and responsibility between management and workers whereas in the past almost all the work and the greater part of the responsibility were thrown upon the men.

Taylor gave birth to the Taylor System or movement which was shared by other important pioneers including Henry Gantt, Frank and Lillian Gilbreth, Harrington Emerson, Louis D. Brandeis, L.P.Alford, A. Hamilton Church and others. Scientific Management brought forth a new concept of attacking management problems with analytical or scientific methods instead of "rule-of-thumb" methods. Industrial engineers and efficiency experts versed in the Taylor System ideas and principles applied scientific management in industry and administrative areas with the goal of increasing human skills as well as mechanical power.

Although Business Administration appears to be one of the last of the main areas of human endeavor to apply statistical methods in the scientific sense, it certainly has made up for lost time during the past thirty years. The first attack was on problems concerning the control of the quality of manufactured products. We all know the modest beginnings of statistical quality control for a period of fifteen years after its discovery in 1924 by Dr. Walter A. Shewhart. We know how difficult it has been to educate business administrators (management), in the value of these newer uses of the old statistical principles. Perhaps there has never been a more intensive educational program in any field than the one started by the invention of the Shewhart control chart.

The first book in the field was by Dr. Shewhart himself in 1931, (4) although it was preceded in 1928 by a book in which T. C. Fry devoted two or three pages to control charts. (5) In 1933 the American Society for Testing Materials produced the well-known "Manual on Presentation of Data" (6) - a little book of 38 pages which was followed by a reprint with a supplement including the "control chart" in 1937. There have been numerous printings of the manual in the years following its introduction.

The British Standards Institution brought out a manual by E.S. Pearson in 1935. (7) This was the first publication of the Institution on the subject of statistical quality control methods. It was followed by a revision in 1942 by B. P. Dudding and W. J. Jennett. (8)

The famous booklet of the American Standards Association known as American War Standards (9) (10) served as principal references during the intensive quality control training program of the World War II, under the direction of the Office of Production Research and Development of the War Production Board. These books and manuals together with a popular book by Leslie E. Simon in 1941 (11) were the fundamental literature of thousands of persons from industrial establishments and colleges who received training in the 8-day courses of that early period of SQC history.

Although statistical quality control and acceptance sampling techniques for the inspection of incoming materials took off to a slow start from 1924 to 1942, the momentum gained during the war period produced a

veritable atomic explosion of ideas and techniques for application of
the theory of probability and sampling to production control once the war
was over. The literature increased until no man alone could keep up with
it all. (12) In many hundreds, even thousands, of periodicals, books,
brochures, and company manuals, one may come upon explanations of appli-
cations of some phase of statistical theory or methodology for use in
industry. This statistical education has spread throughout the world
until today it is a backward industrial country indeed where men are not
making use of these scientific sampling methods. (13)

One of the first papers appearing on the use of statistical quality
control methods in the field of administrative applications was delivered
before the Royal Statistical Society of London by W.F.Newland and E.E.
Neal in 1939. (14) There had been many papers and a few books published
during the fifteen years after the invention of the control chart but
whether they were in the field of textiles, ceramics, steel, automobiles,
electrical or mechanical engineering, or some other area, they centered
on the absorbing subjects of theory and methods in the economic quality
control of manufactured product.

One of the earliest articles devoted to administrative applications
of statistical methods was by W. Edwards Deming and Leon Goefrey in
1941. (15) Although it is a natural step from the use of the control
chart in mass production of manufactured products in the factory to the
use of such charts in mass production of the products of strictly human
operations, someone must be the first to try an adaptation and show that
it is feasible. For those who may not have heard of this first important
use of the control chart in control of human operations, it may be stated
that the Bureau of the Census was reported to have saved approximately
$263,000 through the use of sampling methods and control chart techniques
in the inspection of card-punching and verifying operations connected
with the 1940 census. Something like 51 million cards were sampled out
of over 175 million cards key-punched. Percent defective control charts
were kept for about 475 clerks who qualified for inspection on a sampling
basis. The article gives details of how the program of verification of
clerical work was carried out.

A few years ago the use of statistical methods in the control of
manufactured product was unthought of. Today it is just as sensible to
control the quality of personnel product from the point of hiring,
through daily work, to the finished product of services rendered. A
service organization might well take as its definition of quality con-
trol:

   "Quality control is a system for co-ordinating the quality main-
   tenance and quality improvement efforts of the various groups in
   the organization so as to enable production at the most economic
   levels which allow for full customer satisfaction." (16)

We may hesitate to apply such controls because it is difficult to
deal with "machines" that can talk back, become balky or sulky, cause
work stoppage or delay in production, or even walk off the job. Because
the application is difficult does not justify avoiding it for that is but
an excuse and not a reason. Man can do practically anything he really
wants to do. Capability studies have been made in a rough sort of way
from the hiring of the first hired hand, or rather from the first pur-
chase of the first slave, but scientific capability studies are in their

37

infancy. We lack knowledge of proper standards, particularly standards
of mental work. Though a good deal of progress has been made, we lack
adequate means of readily testing in advance whether a human being has
the capability to turn out a required quantity of mental and/or physical
work to a definite degree of accuracy and consistency required. It
might be said, we are also short in our ability to set up standards of
mental work. The adoption of administrative quality control must wait
on the installation of a system in an office.

It was 1942 before the necessity of reducing waste in labor, mater-
ials, and time of producing quality products gave a real impetus that
generated the statistical quality control movement. But it was not long
before the wide potentialities of the usefulness of SQC could be seen to
extend to business administration or administrative applications as well
as to the industrial areas where it originated.

Credit must be given to W. B. Rice for early articles on the use of
statistical quality control methods in business administration. In 1943
he discussed the adaptation of quality control techniques to office and
financial management problems. (17) He gave a practical example of how
office overtime might be controlled by the use of control charts. The
use of control charts in the office as well as in the factory was advo-
cated.

Another early article by Rice followed in 1944.(18) The author
dealt with the use of control charts in relation to cost analysis. He
also discussed the control of certain expenses with warning of changes in
the nature of the population samples. These articles by Rice and others
about this time appear to mark the beginning of a real interest in the
application of probability methods to business administration problems.
It was not until 1950, however, that a book devoted exclusively to the
use of probability statistics by management was written. (19) This was
26 years after the invention of the control chart. This book by R. K.
Mueller is intended to spur the imagination of management personnel to
consider using statistical control in their business, if it has not al-
ready been used, and, if it has been started as statistical quality con-
trol alone, to extend applications of these methods into more adminis-
trative phases. This is not a book on statistical techniques but rather
one which presents an overall survey of the applications of probability
statistical techniques from a practical management viewpoint. It tells
how to use a statistical control program in the organisation and gives
helps for introducing the technique, for selling it to top management,
supervision, employees, suppliers, and customers, and for administering
the program once it is started.

It is most interesting to follow the development of a new idea such
as the use of statistical methods in industry to see how long it takes
for it to take hold. In this field there was very slow development until
the war stimulated statistical methodology in industry, but in 1945 it
seemed as though the dam of secrecy and occupation with war work had
broken and men became vocal and literary. The flood of articles, papers,
manuals, books spread over the world.

Out of World War II there developed the application of the scientif-
ic method to the over-all operations in hand. The problems of the de-
fense of London from air attack and the urgent problems connected with
submarine warfare and bombing brought forth the application of teams of

brain workers to attack problems as a whole. This was called "Operations Research", or "Operational Research". Operations Research was carried over into peacetime as the application of the scientific method to the overall operations of a business. Operations Research has a broader meaning than plant operation or sales department operation. It looks at problems from the viewpoint of economic value - total cost, profit, and general effectiveness of a program. Operations Research is not statistics (history), it is an experimental science, with the goal of active control of future operations before it. In the application of Operations Research extensive use is made of the mathematics of probability and statistical methods of analysis, although higher mathematics is not necessarily required to solve Operations Research problems. Sampling, analysis of variance and covariance, control charts, significance tests, are valuable tools when needed in this work. Sometimes higher forms of mathematics are required and we note the use of the Waiting line theory, the Search theory, the Game theory, the Communications theory, the Monte Carlo method, or some type of mathematical programming such as linear programming or dynamic programming. In today's forward rush toward scientific business administration techniques there has been a tremendous development in the use of mathematics and electronic data processing machines. Take linear programming as an example of research in this area. Linear programming is a method of optimizing a linear function when the variables are subject to a set of linear constraints. Such a mathematical method might be useful in the administration of scheduling production and shipments, for instance in a transportation problem concerned with the manufacture of a product in a half-dozen plants and distribution to sixty warehouses scattered throughout the country; or the problem of finding the best order for traveling salesmen to tour a number of cities so as to minimize the total distance covered. According to Carl E. Noble (20) - the graphical method works for two independent variables but an algebraic method, called the Simplex Method is used for solving problems involving from 4 to 20 or more independent variables. Thus, it may be necessary to use high speed computers to solve the problem. There are countless statistical methods being evolved by research workers for eliminating arithmetical computations, some of which are discussed by W.R. Vogel. (21)

Although interest has been growing rapidly in the potentialities of mathematical programming techniques in decision making, and in the past six years an extensive literature has been building up as indicated in an article by G.B. Dantzig, (22) it is to be stated that as yet linear programming is not widely used in industry, though there have been a number of very successful applications of the method thus far.

In the newer school of scientific approach to administrative problems the business scientist will:

1) Develop a mathematical model or quantitative hypothesis for the activity to be predicted.

2) Compare the deductions derived by logical manipulations of the abstract model against adequate observational data in controlled experiments to find agreement between abstraction and reality.

3) Conduct critical tests of the overall theory in which predicted substantial effects are observed.

The business scientist will make use of results obtained and re-
ported by earlier writers on similar problems elsewhere. That is one
reason why searching the literature is an important primary step in
successful scientific work. It enables one to learn of the successes
and failures (although failures are not as readily reported as successes)
of others, the problems faced, conditions under which methods were
applied, and the techniques and methods used.

To get an educational toe-hold in this field of study one might read
"Introduction to Operations Research" by C.W.Churchman et al. (23) and
meanwhile read as "warm-up" material the papers by W.E.Alberts (24) and
Harlan D. Mills. (25)

For a person who wishes to learn something about linear programming
without running head-on into higher mathematics, I suggest three short
articles ahead of those mentioned under (20), (21), and (22). These are:
H.T. Schwan (26), A.W.Tucker (27), and A.C.Rosander (28). These last
three articles will be found in the March 1956 issue of Industrial
Quality Control.

The need for more than the usual high school and college allotments
of mathematics in order to understand the discoveries in statistical
methods research in the field of business administration, is nowhere more
evident than in such articles as "The Inventory Problem" by J. Laderman
et al., (29) and "The Use of Mathematics in Production and Inventory
Control" by A. Vassoni. (30)   There are many articles in this area, a
few of which are given in references(31), (32), (33), (34), (35), and
(36).

"It is becoming increasingly apparent," Mr. Vassoni states, "that
some new scientific and mathematical theories are now being developed in
this country which, in combination with the remarkable performance of
electronic data processing machines, will bring about significant changes
in current managerial techniques." In the course of studying production
and inventory control, Mr. Vassoni reports it was repeatedly necessary to
explain his mathematical theory to operating personnel, many of whom have
had little or no formal training in mathematics at the college level.
One author, T.M.Whitin, (36) says, "The next few years should bring about
much additional research which will help decide to what extent industrial
application of the various types of inventory control systems will be
profitable. Contrary to the situation in the natural sciences, univer-
sity research in the field of inventory control is far ahead of current
practice."

Unless I am completely incapable of diagnosing the present situa-
tion, it will be several years before any but a few colleges of business
administration will have courses of training in management, accounting,
marketing, production, or any other major field in which articles such
as those cited pertaining to the inventory problem will be assigned as
required reading. The perusal of any college catalogue will disclose
that students in colleges of business administration are not required to
have the mathematical background to enable them to cope with the advanced
and ever advancing literature dealing with such problems. And neither,
I might hazard an opinion, are present day business administration
faculties required to have such training. These startling words may be
news to those not in close touch with the situation, but I believe them
to be a statement of the truth of the present day college situation.

Colleges of Business Administration do not, in general, require any mathematics beyond high school algebra and geometry as prerequisite to courses. These high school subjects are almost entirely forgotten, and skill in them reduced to zero, by the time students enter courses in college in which such subjects might be of some use. In this country college mathematics has been considered as unnecessary to the education of a future businessman as a foreign language. A quarter or semester course in elementary statistics is generally required of all business administration students, a course which is taken with some fear and trembling and much apologising for a lack of mathematical aptitude. Students can all read and write after a fashion but for the most part are grossly inept, if not downright illiterate mathematically. It will not be possible for the brilliant research being done to be taught in undergraduate courses in colleges until the entire educational system is changed to include much more training in mathematics in the entire twelve-year period leading up to college. The fundamental place to start this radical change is in the home. There will need to be considerable education of parents to encourage the study of mathematics by their children, and to insist that teachers all along the line work their children harder and longer. At present the teaching of courses dealing with mathematics in any form other than simple grocery-store arithmetic is as painful as building a house on a foundation too weak to support the superstructure.

I am not arguing that higher research into the use of statistical methods in business administration or administrative methods should slow down. Far from it, for that would be folly, indeed. I simply want to put in a strong word for more mathematics and statistics in the curricula of grade schools, high schools, and colleges. Mathematics is the language of numbers. However, we treat mathematics in this country as though it were a language we never expected the businessman to make continuous use of in the future. We do not learn enough of it over a long enough period of time. It may be that people in general have not realised that the world of business and industry has changed and now requires greater mathematical understanding for real progress. It may be because of the overcrowded conditions in schools due to the current philosophy that every child is entitled to a high school education, and should be in school until 18 years of age, that we find mathematics being reduced in importance for the sake of the mentally weak. Because mathematics is not entertaining like TV and fun like tossing a basket-ball around, we introduce many softer courses of instruction adopting new philosophies to rationalize our actions. Thus we may advocate little or no homework and permit the students to sidestep courses dealing with data, formulae, and computations. Perhaps what we need in this country is an entirely new method of teaching arithmetic. Arithmetic might be considered easy in time, as easy as English had one been exposed in youth to the method taught at the Trachtenberg Institut in Zurich, Switzerland. (37) It is little wonder that it becomes more and more difficult or impossible for the college student and the businessman to read advanced articles in the trade and professional journals of the day; they are just not equipped to do it.

A much more down-to-earth and readily understood article on inventory control, that is, one that would be more likely to be read and used in practice, and find a more receptive audience today than those listed in references (29) to (36), is one by W.F.Hoehing. (38) Although this article applies the theory of probability as exemplified in the

41

cumulative Poisson distribution, the mathematics are not beyond a second
or third course in Business Statistics, and thus there is a reasonable
chance that a few of the current business administration students might
learn about it, understand it, and eventually use the method. Even more
readily understood articles pertaining to inventory control and sales
forecasting are those by Frances B. Newman (39) and Charles A. Bick-
ing. (40) The latter article shows how the control chart can be used
to assist management obtain internal control and gives examples covering
1) merit rating of production personnel, 2) trend in manufacturing cost,
3) industrial lost-time injuries, and 4) forecast of sales volume. In
another article (41) Bicking strongly supports the adaptation of statis-
tical methods to administrative applications and managerial problems.
Here he brings out the applications pertaining to inventory control in-
volving twelve products distributed from four warehousing points; an
application pertaining to the rating of technical personnel; and one in
the analysis of indirect expense as a percent of sales. Another "easy-
to-read" article in the same vein is by Guy G. Parkin. (42) These last
five articles are not likely to discourage the average business magazine
reader if he has had an introduction to statistical theory and the
ambition to learn new methods of dealing with administrative problems.

Mathematical research development in marketing may have had its in-
spiration in such early papers as the one by W. Edwards Deming in
1941. (43). Here was discussed the need for directing the collection of
data for marketing research towards the identification of causes of in-
creases or decreases in sales. Benefits to be derived through the
application of Shewhart methods of quality control were discussed with
features of the plan. The part sampling plays in marketing research was
presented with an outline of the main aspects of the statistician's job.
Also, an original outline was given of two principal types of statistical
problems and the relationship between the evidence, degree of belief in
the prediction based on the evidence, the action to be taken, and the
possible consequences of the action.

An early milepost along the road of application of scientific sam-
pling in marketing research was by Theodore H. Brown, in 1942. (44)

One comes across an occasional article utilizing scientific sam-
pling in the field of marketing such as M.E. Brunk and W.T. Federer. (45)
This article pertaining to research designed to measure consumer wants of
a commodity and reporting on some 34 experiments and results thereof
indicates the direction statistical research is taking in this area.
It is evident that the universities are not developing research in the
marketing area as rapidly as in production and inventory control accord-
ing to M. Anshen. (46) Anshen states "in this day marketing budgets
involve commitments on too large a scale to permit executives to con-
tinue to accept historical patterns in a management world in which
rational exploration of alternative strategies and planned optimizing are
becoming standard operating procedures." These words indicate that per-
haps we are on the verge of a marketing literature that will include
rules of general applicability or a method of research that will provide
answers to such common questions as:

1) How much money should we spend on advertising next year?

2) What is the best division of the total advertising budget among
the various available media?

3) How can we make rational decisions with respect to complementary budgets for advertising and other sales activities in local areas?

Today the rules for making decisions in setting advertising budgets come under the heading of:

1) A fixed percentage of last year's sales (or this year's sales forecast, the percentage being determined by past practices or the pattern in the industry;

2) A fixed number of dollars per unit sold last year, or forecast to be sold this year;

3) A fixed-percentage rule modified by recent profit performance and short-term profit forecasts.

The reason for this situation is lack of knowledge on how to establish direct cause-effect relationships between sales promotion outlays and sales to ultimate consumers. The above three questions do not appear to be answerable through existing methods. However, with all of the writing on the advantages of scientific methods many marketing research men still go no further in their analyses than the use of relative comparisons in percentage, using methods from which it is impossible to measure the error of their sampling results.

Work measurement is an area in which statistical methods have been developing. This field is closely allied with industrial engineering, as are statistical quality control, operations research, and cost accounting. Work measurement means a complete study and description of an operation including selecting appropriate characteristics of the operation to be time-measured, improving the methods and procedures where possible, setting up a control system for scheduling, and in the case of incentive standards, controlling pay whereby countable output can be related to time input. In statistical quality control, experimental design, and operations research, powerful techniques have been developed that can be used in "work measurement". For one who wishes to learn more about this area of statistical methods in business administration, reference should be had to the articles by C.L.Brisley, (47)(48) and then one or two books such as "Work Sampling" by R.M.Barnes (49) or "Work Measurement" by A. Abruzzi. (50) A more recent article, that brings out the usefulness of various mathematical methods such as queueing models, linear programming, and multiple regression, is by J. M. Allderidge.(51)

Statistical methods reach into the business administration of various types of endeavor such as department store operation. An easily understood article by C.S.Brinegar (52) brings out many interesting aspects readily applicable. Fred Streit (53) outlines the program used in establishing and maintaining quality standards in a mail order business. For students of sampling, it is interesting to compare discussion of sampling in the public utility field in "Some Theory of Sampling" by W. Edwards Deming, (54) and the article "Statistical Methods for Appraising Public Utility Property," by E.T.Magruder. (55) These show advances being made in the practical use of scientific sampling design in business administration. A utilization of the actuarial application of statistical information for maintenance of airplane engines predicts the far-reaching uses of probability statistics. Methods that are working daily to the advantage of business are illustrated by

Allan M. Hull and W.W.Wilcox. (56)(57)

We come to two principal areas of use of statistical methods in business administration which are developing rapidly and will continue to expand until they are common to all business. I refer to the use of sampling methods in auditing and the use of statistical methods in clerical work of all kinds, including accounting.

It is fascinating to follow a field of endeavor such as accounting from its inception in the dim past when temple records were kept on clay tablets with a stylus in cuneiform characters down through the centuries to the modern application of scientific methods and high speed electric machinery. We learn that double-entry bookkeeping had its origin in the 11th or 12th century as far as anyone knows, and the first printed word on this subject still in existence is in a book by a Franciscan monk in 1494. (58)(59) But it was not until after the publication of "Auditing" by L.R.Dicksee in 1892 that auditing began to take on the modern aspect of a profession. (60) What appears to have been the first approach to the use of probability sampling in auditing work is the article "The Efficacy of Tests," by Leo A. Carman. (61) Carman presented a table which seeks the mathematical determination of the relative degree of probability of discovering one out of a specific number of fraudulent or false items included in a larger group, by the examination of samples of various sizes. The underlying idea was that when a single false item is disclosed a discovery of sufficient value has been made to justify a completely new consideration of the sampling process with the possible alternative of a 100 percent check.

Let us see how this pioneering approach of Carman in 1933 was carried forward in succeeding articles and books. R.H.Prytherch in 1942 discussing the problem of test checking states, "Judgment must be used to determine the extent to which various parts of the work must be checked and internal control is a factor in these decisions." He outlines the statistical rule as to the reliability of samples and its application to test checking reproducing the table presented by Carman nine years previously. (62)

The same Carman table is mentioned again (p. 21) in 1950 by L.L. Vance in his book "Scientific Method in Auditing." (63) This book was a pioneering effort to introduce probability sampling in a field which has been using judgment sampling methods exclusively from the inception of auditing as a profession. Again, Carman's table was mentioned by Howard L. Jones in 1952, and in 1953. (64) Once more this table is brought up by Vance and Neter (pages 100-101) in 1956. (65) This fabulous table bobbed up again in an article by H. Arkin in July 1957, wherein Table III is credited to Carman with his idea of using the theory of probability in auditing work. (66) Evidently the idea of scientific sampling can no longer be ignored in the field of accounting and auditing, though it is going to be a long struggle to convince the old school of accounting practitioners of the practical efficacy of these new techniques. The auditor must always be prepared to interpret the results of sampling and decide upon consequent action. His judgment will always be required as an expert in the subject matter to state his problem with exactness, decide what is a suitable frame for sampling, what the frame is, and how to obtain the information from any unit of the frame. This is not the province of the statistician although he might be of help in the decisions. To quote from the article by Arkin just

mentioned:

> "In general, the accounting profession has not fully realized that
> during the past 50 years a body of knowledge generally referred to
> as sampling theory and practice has been developed on a scientific
> basis. This sampling knowledge has been widely applied in all
> fields. Not only has it found wide acceptance and application in
> the sciences, but in industry and business as well. Since auditing
> is looked upon as an art and not a science, auditors have not
> searched out these methods."

Following up the first publication in 1950, (63) another book of
larger proportions, namely, "Statistical Sampling for Auditors and
Accountants" was published by L.L.Vance and John Neter in 1956. (65)
This is truly a pioneering venture to introduce scientific methods into
the realm where judgment sampling has always held complete and unob-
structed sway. (See book review in The Accounting Review by J.L.
Roth.) (67) Today it is becoming common to find articles dealing with
statistical procedures and probability sampling in auditing and account-
ing work. One such article by A.W.Patrick (68) briefly explains the use
of the control chart idea and advocates the expansion of cost accounting
to include interpreting the significance of variations. An important
basic article addressed to accountants advocating the exploration of the
possibilities of adapting statistical techniques to accounting problems
is by R.M.Trueblood. (69)

Two fine papers worth reading by anyone interested in the applica-
tion of statistical methods in auditing and accounting are those by
A.C.Smith (70) and Herbert Arkin. (66) Mr. Arkin states reasons for
failure of statistical sampling method to permeate the field of auditing
more quickly; they stem from the fact that the auditor's sampling problem
involves the fact that the auditor does not place sole reliance upon or
base his final decisions exclusively upon the results of a single test.
The evidence obtained from each test is evaluated together with the re-
sults of other tests and other types of examinations and explorations.
In this respect the auditor's sampling operation differs greatly from
that encountered in other fields. Sampling techniques must be designed
to meet the specific need of the auditor.

The list of articles elucidating statistical methods in auditing and
accounting control is growing with every issue of The Journal of Account-
ancy, The Accounting Review, The Journal of the American Statistical
Association, Industrial Quality Control, the New York Certified Public
Accountant, the National Association of Cost Accountants Bulletin, to
mention a few journals. Rather than list many articles it is suggested
that the interested researcher should cover the indexes of these six
journals for the past seven or eight years and also scan the indexes of
the last half-dozen ASQC Conference Transactions, and thus identify
articles in this area. By keeping up with the current literature in
these important publications, one can watch the progressive development
of the adaptation of statistical methods to auditing and accounting.

To get a start in the right direction, it is suggested for those
who are more interested in the application of statistical methods to
clerical work control and accounting procedures than they are to audit-
ing, that the following articles be read: W.F.Buhl, (71); Howard L.
Jones, (72)(73); and A.C.Rosander, (74)(75). In these latter papers

Mr. Rosander describes the use of probability statistics in the confirmation of accounts receivables, the aging of accounts receivables, the control of payrolls, the control of inventory, and the control of costs. He discusses sampling in the estimation of the condition of physical assets, in settling inter-company accounts, and work sampling as a basis for cost analysis.

Regarding cost accounting and cost control there are several important articles readily at hand: Carl E. Nobel (76)(77)(78) and E. W. Gaynor. (79) There are, of course, other articles but these are good as an introduction to the area.

One could not complete even such a brief recital of the literature as this without mention of several additional individuals who appear as pioneers in the application of statistical methods in business administration, with particular reference to administrative applications. Dale L. Lobsinger, a past-president of the American Society for Quality Control, and first chairman of the Administrative Applications Division, is well known for his work in quality control for the air lines. An article such as "Some Administrative Attributes of Statistical Quality Control," (80) should open the eyes of many persons to the positive benefits that have accrued to management through the use of these techniques in gaging production efficiency and accurately determining facility capacity under fixed conditions. Some of the areas in which statistical quality control techniques are used in the air lines are: accounting work, ticket sales, space control, food ordering, aircraft delays, air freight billing, aircraft maintenance, inventory control, and interline billing. In regard to the last-named type of work, reference should be had to W.G.Dalleck. (81)(82) These articles describe how modern scientific methods have been adopted by the airlines (and railroads too) thus cutting clerical costs for the benefit of owners, management, and the public.

And finally, a few papers dealing with statistical control of clerical operations may give the reader the courage to attempt application of these modern methods in his own office. I shall merely list the names of the authors and reference may be had to the articles in the bibliography:

J. M. Ballowe (83), B.B.Murdock (84), E.W.Gardiner (85), R.B. Shartle (86)(87), Ruth A. Hanna (88), A.L.Hart (89), R.B.Selover (90), D.J.Blackwell (91), and A.C.Rosander (92).

It would be possible to list many more articles pertaining to the application of statistical methods for the control of clerical operations, quality control in the office, for the control of large scale invoicing, checking vendors' invoices, checking tax returns, improvement of clerical work quality, and so forth. I apologize to everyone who has contributed to this important field and whose paper has been omitted. The papers included and the importance of the companies represented by the authors ought to convince any present or prospective businessman or educator that we have not only come a long way from the ancients in our record keeping but have devised wonderful methods for analyzing the results of our efforts. Business Administration is well on its way in the application of scientific methods in decision making and no man can say definitely how rapidly it will develop in the near future. It would certainly seem that the American Society for Quality Control can play

a very important role in the stimulation of the study of mathematics and science not only by the young in all levels of school education but by adults who wish to keep abreast of the developments of this scientific and electronic age.

References

1. Proceedings of Symposium on Design of Industrial Experiments, Institute of Statistics, University of North Carolina, Raleigh, 1956.

2. Savage, J.R., "Bibliography of Non-parametric Statistics and Related Topics," Journal of the American Statistical Association, 48, Dec. 1953: 855-906.

3. Taylor, Frederick W., The Principles of Scientific Management, Harper Bros., New York and London, 1911.

4. Shewhart, Walter A., Economic Quality Control of Manufactured Product, D. Van Nostrand Co., New York, 1931.

5. Fry, T.C., Probability and its Engineering Uses, D. Van Nostrand Co., New York, 1928.

6. Manual on Presentation of Data, American Society for Testing Materials, Philadelphia, 1933. Also, 1940, 1941, 1945, etc.

7. Pearson, E.S., The Application of Statistical Methods to Industrial Standardization and Quality Control, British Standards Institution, London, 1935.

8. Dudding, B.P. and Jennett, W.J., Quality Control Charts, British Standards Institution, London, 1942.

9. Guide for Quality Control and Control Chart Method of Analyzing Data, American War Standards Z1.1-1941 and Z1.2-1941, American Standards Association, New York, 1941.

10. Control Chart Method of Controlling Quality During Production, American War Standards Z1.3-1942, American Standards Association, New York, 1942.

11. Simon, Leslie E., An Engineers' Manual of Statistical Methods, John Wiley and Sons, New York, 1941.

12. Butterbaugh, Grant I., A Bibliography of Statistical Quality Control, University of Washington Press, Seattle, 1946.

12. Butterbaugh, Grant I., A Bibliography of Statistical Quality Control-
a. Supplement, University of Washington Press, Seattle, 1951.

13. International Journal of Abstracts on Statistical Methods in Industry, International Statistical Institute, The Hague,   Vols.: I, 1954; II, 1955; III, 1956; IV, 1957; V, 1958.

14. Newland, W.F. and Neal, E.E., "Statistical Control of the Quality of Telephone Service," Journal of the Royal Statistical Society, Supplement 6, 1939: 25-50.

15. Deming, W.E. and Goefrey, L., "On Sample Inspection in the Processing of Census Returns," *Journal of the American Statistical Association*, 36, Sept. 1941: 351-360.

16. McCone, Robert. (United Air Lines, Seattle). From correspondence Nov. 29, 1957.

17. Rice, W.B., "Quality Control Applied to Business Administration," *Journal of the American Statistical Association*, 38, June 1943: 228-232.

18. Rice, W.B., "Statistical Uses of Accounting Data," *Accounting Review*, 19, July 1944: 260-266.

19. Mueller, R.K., *Effective Management through Probability Control*, Funk and Wagnalls, New York, 1950.

20. Noble, Carl E., "The Simplex Method for Solving Linear Programming Problems, *Industrial Quality Control*, 13, March 1957: 5-9.

21. Vogel, W.R., "Mathematical Linear Programming," *Proceedings of the 11th Midwest Quality Control Conference*, Minneapolis, Oct. 1956: 101-109.

22. Dantzig, G.B., "Recent Advances in Linear Programming," *Management Science*, 2, Jan. 1956: 131-144.

23. Churchman, C.W.; Ackoff, R.L.; and Arnoff, E.L., *Introduction to Operations Research*, John Wiley and Sons, New York, 1956.

24. Alberts, W.E., "Quality Control, Industrial Engineering, and Operations Research," *National Convention Transactions 1955*, American Society for Quality Control: 177-182; also *Industrial Quality Control*, 12, May 1956: 29-31.

25. Mills, H.D., "Integrating Operations Research into a Business," *National Convention Transactions 1955*, American Society for Quality Control: 397-403.

26. Schwan, H.T., "Practical Linear Programming Applications," *Industrial Quality Control*, 12, Mar. 1956: 4-8.

27. Tucker, A.W., "Linear Programming," *Industrial Quality Control*, 12, Mar. 1956: 8-11.

28. Rosander, A.C., "The Use of Linear Programming to Improve the Quality of Decisions," *Industrial Quality Control*, 12, Mar. 1956: 11-16; 17 references.

29. Laderman, J.; Littauer, S.B.; and Weiss, L., "The Inventory Problem," *Journal of the American Statistical Association*, 48, Dec. 1953: 717-732.

30. Vazsoni, A., "The Use of Mathematics in Production and Inventory Control," *Management Science*, Part I, Oct. 1954: 75-85; Part II, April-July 1955: 207-223.

31. Dvoretzky, A. et al., "The Inventory Problem I, II," <u>Econometrica</u>, <u>20</u>, April–July 1952: 187–222.

32. Feeney, G.J., "A Basis for Strategic Decisions on Inventory Control Operations," <u>Management Science</u>, 2, Oct. 1955: 69–82.

33. Bellman, R. et al., "On the Optimal Inventory Equation," <u>Management Science</u>, <u>2</u>, Oct. 1955: 83–104.

34. Beckman, M. and Muth, R., "An Inventory Policy for a Case of Lagged Delivery," <u>Management Science</u>, 2, Jan. 1956: 145–155.

35. Simon, H.A. et al., "Controlling Inventory and Production in the Face of Uncertain Sales," <u>National Convention Transactions 1956</u>, American Society for Quality Control: 371–378.

36. Whitin, T.M., "Inventory Control Research: A Survey," <u>Management Science</u>, <u>1</u>, Oct. 1954: 32–40; 17 references.

37. <u>Trachtenberg German Language Correspondence Course</u>, Institut Trachtenberg, Zurich, Switzerland.

38. Hoehing, W.F., "Statistical Inventory Control," <u>Industrial Quality Control</u>, <u>13</u>, Jan. 1957: 7–13; also <u>National Convention Transactions 1955</u>, American Society for Quality Control: 385–395.

39. Newman, Frances B., "A Technique for Adjusting Sales Predictions to Actual Sales Trends: a Method of Inventory Control," <u>Industrial Quality Control</u>, <u>7</u>, May 1951: 29–31.

40. Bicking, Charles A., "The Application of Quality Control to Administrative Problems," <u>Industrial Quality Control</u>, <u>6</u>, May 1950: 21–25.

41. Bicking, Charles A., "Quality Control as an Administrative Aid," <u>National Convention Transactions 1955</u>, American Society for Quality Control: 347–357.

42. Parkin, Guy G., "Statistical Technique for Forecasting Sales," <u>Industrial Quality Control</u>, <u>12</u>, Nov. 1955: 5–12; also <u>National Convention Transactions 1954</u>, American Society for Quality Control: 85–91.

43. Deming, W. Edwards, "On Some Contributions of Statistical Inference to Marketing Research," <u>Ohio Conference of Statisticians on Business Research</u>, 1941: 13–24.

44. Brown, Theodore H., "Scientific Sampling in Business," <u>Harvard Business Review</u>, <u>20</u>, Apr. 1942: 358–368.

45. Brunk, M.E., and Federer, W.T., "Experimental Designs and Probability Sampling in Marketing Research," <u>Journal of the American Statistical Association</u>, <u>48</u>, Sept. 1953: 440–452.

46. Anshen, M., "Management Science in Marketing: Status and Prospects," <u>Management Science</u>, <u>2</u>, Apr. 1956: 222–231.

47. Brisley, C.L., "Work Sampling (Ratio Delay Study) Techniques," Quality Control Conference Papers 1955, American Society for Quality Control: 115-125.

48. Brisley, C.L., "How You Can Put Work Sampling to Work," Factory Management and Maintenance, July 1952: 83-89.

49. Barnes, R. M., Work Sampling, University of California at Los Angeles, Los Angeles, 1956.

50. Abruzzi, A., Work Measurement, Columbia University Press, New York, 1952.

51. Allderidge, J.M., "Designing the Measurement and Analysis of Work," National Convention Transactions 1957, American Society for Quality Control: 191-203; 13 references.

52. Brinegar, C.S., "Department Store Uses of Statistical Quality Control," National Convention Transactions 1954, American Society for Quality Control: 311-315; and Industrial Quality Control, 12, Aug. 1955: 3-5.

53. Streit, Fred, "Mail Order Acceptance Sampling," Industrial Quality Control, 13, July 1956: 15-18.

54. Deming, W. Edwards, Some Theory of Sampling, John Wiley and Sons, New York, 1950: p. 95.

55. Magruder, E.T., "Statistical Methods for Appraising Public Utility Property," National Convention Transactions 1954, American Society for Quality Control: 27-41.

56. Hull, A.M. and Wilcox, W.W., "Statistical Quality Control Applications for Improving Aircraft and Engine Maintenance," National Convention Transactions 1954, American Society for Quality Control: 333-360.

57. Hull, A.M., "Optimizing Maintenance Costs through Quality Control," Industrial Quality Control, 12, June 1956: 4-8.

58. Geijsbeek, J.B., Ancient Double-Entry Bookkeeping, John B. Geijsbeek, Denver, Colorado, 1914. An English translation of a German translation from Summa de Arithmetica Geometria Proportioni e Proportionalita by Lucas Pacioli, 1494. Double-entry bookkeeping makes up one section in Pacioli's book on mathematics.

59. Butterbaugh, Grant I., "Dr. Stands for Debit," The Accounting Review, July 1945: 340-344.

60. Dicksee, L.R., Auditing, Gee and Co., London, 1892.

61. Carman, Leo A., "The Efficacy of Tests," The American Accountant, 18, Dec. 1933: 360-366.

62. Prytherch, R.H., "How Much Test Checking is Enough?" Journal of Accountancy, 74, Dec. 1942: 525-530.

63. Vance, L.L., Scientific Method for Auditing, University of California Press, Berkeley, 1950.

64. Jones, H.L., "Some Problems in Sampling Accounting Records," Sixth Annual Conference Papers, 1952, American Society for Quality Control: 209-250; also Industrial Quality Control, 10, Nov. 1953: 6-13.

65. Vance, L. L. and Neter, John, Statistical Sampling for Auditors and Accountants, John Wiley and Sons, New York, 1956.

66. Arkin, H., "Statistical Sampling in Auditing," New York Certified Public Accountant, 37, July 1957: 454-469.

67. Roth, J.L., "Review of Statistical Sampling for Auditors and Accountants by L.L.Vance and J.Neter," The Accounting Review, 32, Jan. 1957: 169-170.

68. Patrick, A.W., "A Proposal for Determining the Significance of Variations from Standard," The Accounting Review, 32, Oct. 1957: 587-592.

69. Trueblood, R.M., "The Use of Statistics in Accounting Control," New York Certified Public Accountant, 23, Oct. 1953: 619-626; and National Association of Cost Accountants Bulletin, 34, July 1953: 1561-1571.

70. Smith, A.C., "The Accounting Profession's Growing Interest in Statistical Sampling," New York Certified Public Accountant, 27, July 1957: 451-453.

71. Buhl, W.F., "Statistical Controls Applied to Clerical and Accounting Procedures," National Convention Transactions 1955, American Society for Quality Control: 9-25.

72. Jones, Howard L., "Quality Control Techniques for Inspecting Clerical Work," Industrial Quality Control, 11, May 1955: 42-45.

73. Jones, Howard L., "Some Problems in Sampling Accounting Records," Industrial Quality Control, 10, Nov. 1953: 6-13; also Sixth Annual Conference Papers, 1952, American Society for Quality Control: 209-250.

74. Rosander, A.C., "Probability Statistics in Accounting," Industrial Quality Control, 11, May 1955: 26-31.

75. Rosander, A.C., "Sampling Applied to Accounting Problems," National Convention Transactions 1956, American Society for Quality Control: 283-297.

76. Noble, Carl E., "Statistical Cost Control in the Paper Industry," Industrial Quality Control, 9, May 1953: 42-46; also Sixth Annual Conference Papers, 1952, American Society for Quality Control:9-15.

77. Noble, Carl E., "Cost Accounting Potentials of Statistical Methods," National Association of Cost Accountants Bulletin, 33, Aug. 1952: 1470-1478.

78. Noble, Carl E., "Calculating Control Limits for Cost Control Data," *National Association of Cost Accountants Bulletin*, 35, June 1954: 1309-1317.

79. Gaynor, E.W., "Use of Control Chart in Cost Control," *National Association of Cost Accountants Bulletin*, 35, June 1954: 1301-1309.

80. Lobsinger, Dale L., "Some Administrative Attributes of Statistical Quality Control," *Industrial Quality Control*, 10, May 1954: 20-24.

81. Dalleck, W.G., "Quality Control at Work in Airline Accounting," *National Convention Transactions 1954*, American Society for Quality Control: 489-498.

82. Dalleck, W.G., "Inductive Accounting - Settling Interline Accounts by Sampling Methods," *Industrial Quality Control*, 13, Dec. 1956: 12-16.

83. Ballowe, J.M., "Statistical Quality Control of Clerical and Manual Operations," *Conference Papers: 4th Annual Convention 1950*, American Society for Quality Control: 21.

84. Murdock, B.B., "Some Aspects of Quality Control as Applied to Clerical Operations," *Industrial Quality Control*, 7, May 1951: 39-40, 44.

85. Gardiner, E.W., "Statistical Quality Control in the Office," *Proceedings, Annual Mid-Atlantic Regional Conference*, American Society for Quality Control, Feb. 1951.

86. Shartle, R.B., "How a System of Continuous Scientific Sampling is Used to Control Accuracy in Large Scale Invoicing," *Journal of Accountancy*, 94, Aug. 1952: 167-171.

87. Shartle, R.B., "Quality Control Applied to Clerical Operations," *National Convention Transactions 1956*, American Society for Quality Control: 303-312.

88. Hanna, Ruth A., "Applying Statistics to Checking of Vendors' Invoices," *National Association of Cost Accountants Bulletin*, 35, Dec. 1953: 503-511.

89. Hart, A.L., "Clerical Quality Control has Two Dimensions," *Office Executive*, 1, July 1953: 10.

90. Selover, R.B., "Some Applications of Quality Control Techniques to Clerical Work," *Industrial Quality Control*, 12, Aug. 1955: 5-8; also *Quality Control Convention Papers 1954*, American Society for Quality Control: 147-153.

91. Blackwell, D.J., "Acceptance Sampling - Another Method of Clerical Quality Improvement," *Industrial Quality Control*, 11, June 1955: 17-20.

92. Rosander, A.C., "Probability Statistics Applied to Tax Returns," *Industrial Quality Control*, 9, May 1953: 46-56.

MACHINE CAPABILITY VERSUS PRODUCT TOLERANCE

Edward Oakley
Delco-Remy Division
General Motors Corporation

In recent years more is being said and written about Machine Capabilities. Just what do we mean by "Machine Capability" or "Capability Analysis"? What do we mean by "Process Capability"? And more recently a new term has entered the picture, "Machine Reliability." What is its meaning?

It might be well to make the distinction so that we may have a common understanding of each of the terms.

Let's analyze the meaning of each of the above in layman's words.

1. Process Capability

   The accepted meaning of this is that a processing method which includes the man, machine and material is being studied to determine total variability and process stability, usually termed "state of statistical control." Invariably, time is an important factor to consider because changes in quality level will occur as tools wear or are replaced; or as the man makes corrective adjustments; or when materials are changed; or when any combination of the above three changes significantly.

2. Machine Reliability

   This is a new term which has entered the picture in recent years. Primarily, it has to do with the long term aspects of machine longevity and total expense of maintenance during its total life. Due to the records that must be maintained, this particular study will require close coordination between Quality Control, Maintenance and Production. Properly applied, this study can be very valuable in making comparisons among different makes of machines doing comparable work.

3. Machine Capability Study Or Analysis

   The accepted meaning of this is that a machine or fabricating device is being studied under controlled conditions to ascertain natural or inherent variation. That is, the man is restricted from making any tool adjustments and the quality of the material being processed is known to be of a given quality. Tool wear and other influencing factors are also considered. Primarily we are concerned with what the machine or device can do over a stated period of time and number of pieces of work.

This paper deals with the preliminary planning necessary in making accurate decisions regarding machine capabilities. The intent is to show how a machine capability can be made, using basic S.Q.C. principles, fortified with sound logic, and applied with common sense.

Anyone whose full time is devoted to Statistical Quality Control in a competitive business finds that decisions must be made quickly and

accurately. This is particularly true when studies are being made in a machine builder's plant, or in a manufacturing department. The people concerned with the problem want the answers "now", and not some time in the future. The writer hopes that the experience we have gained in machine capability studies at Delco-Remy Division, General Motors Corporation, will help others in Statistical Quality Control to be more effective in their respective companies.

A case history of an actual problem is presented for your study. The planning steps and problems encountered are listed as a guide to anyone concerned with a similar problem.

Although some discussion will be devoted to the measurement of dispersion, it is felt that anyone attempting a machine capability study should know before hand which method of statistical analysis will be used.

Suffice to say, the more common and most widely used methods of measuring dispersion are:

1. Frequency Distribution (Tally)
2. Calculating Standard Deviation
3. Probability Paper
4. $\bar{X}$ & R Chart
5. $\bar{R}/d_2$
6. Span Plan – by Leonard A. Seder
7. Lot Plot – Dorian Shainan
8. Analysis of Variance

The author prefers to use 1, 3, 4 and 5 above.

We might begin by first considering some of the preliminary steps necessary prior to making a machine capability study.

Step 1

It is understood that the individual assigned to the job should have a background in quality control methods and it is highly desirable to have had experience in manufacturing. In any case, he should be familiar with the processing of the part, because this will influence his decisions in the capability study.

Step 2

(a) Become familiar with the part. Study the dimensional characteristics involved. After some study it will become apparent as to which dimensions are most important to the capability study.
(b) Some dimensions may be checked with a snap gage because they are easy to control. Other dimensions must be checked with gages where readings of variability can be obtained. Usually, these are the dimensions which are used to determine whether the machine being studied is capable of processing a desired part to a given standard.
(c) The design of a part should also be scrutinized for any possible effect on the study.
(d) Consider the kind of material used to make the part. This is frequently a factor in tool life, kind of tooling, rate of

54

speeds and feeds, etc., and will also extend into gaging procedures and gaging accuracy.

Step 3

After becoming familiar with the part, the next step is to check the condition of the pieces coming to the machine. The existing conditions will have a definite bearing on the capability study. Frequently, the roughing operation must be corrected before a capability study of the finish operation can be attempted. Care should be taken that parts are not specifically selected for the study. In all cases, the normal flow of acceptable parts should be used for the study.

Step 4

Study the process under question. What is the fabricating process? (Cut, grind, drill, pierce, hob, shape, form, etc.) Is it a single spindle or multiple spindle, single station or multiple station machine? How are adjustments made? Automatically or manually? How fast must the operation be performed? (Time cycle of processing time.)

Step 5

Verify the accuracy of all gages or instruments to be used in the study. Make several checks with these gages to determine their ability to repeat a reading. Determine definitely whether the proper gages are being used. For instance, snap gage versus micrometer, micrometer versus shadow graph, two (2) orifice air gage versus three (3) orifice air gage, etc.

Step 6

Investigate all sources of information. Contact all people who may be concerned with the problem.

1. Engineering
2. Inspection
3. Production
4. Processing
5. Any others

After the Quality Control man is thoroughly familiar with the part, gages, and process, and has obtained all the available information relative to the part, then he must prepare himself to face the first obstacle in his path. He must be prepared to overcome opposition to a new way of determining a machine's capability. Our experience, at first, was that the machine builders were not too receptive to having extensive studies made in their plants. They felt that all final adjustments made after the machine was installed would make the machine capable of meeting all the prescribed standards.

We found that a negative attitude toward statistical methods stemmed from several preconceived ideas. First, they lacked knowledge of Statistical Quality Control. Second, they felt that Quality Control only applies to manufacturers of small parts. Third, they failed to

55

grasp the idea that as closer tolerances are demanded of us, we, in turn, expect to obtain equipment capable of holding tolerances far more exact than the requirements needed years ago.

As a result of our experience, a few suggestions are being submitted to help overcome these obstacles. These, we will discuss from the standpoint of anticipated purchase of new machines.

1. When entrance is made into the machine builder's plant, find out how much is known about Statistical Quality Control. Our experience indicates that they have heard about it but feel they have no use for it.

2. Be prepared to spend some time explaining S.Q.C. Be certain that all explanations are confined to simple language that can be understood by a layman.

3. Be certain the machine operator understands what you are attempting to do. Be deliberate in explaining Quality Control techniques. Further, let him become acquainted with your gages. Keep in mind that in the development of the machine, they used their own gages which may be different than the ones you intend to use. The operator's cooperation is a must.

4. Explain to all people involved, directly or indirectly, the approach you intend to take and the net results you expect.

5. Have the operator set the machine as near the mean dimension of the part print as possible. Run machine without further adjustments until a sufficient number of pieces have been run for the study.

6. Identify and maintain the order in which the pieces were run. The maintenance of chronological order may be required for further analysis if and when necessary.

7. When the capability analysis is complete, a meeting of all persons concerned should be called. An honest and open-minded discussion should then occur. If the study is satisfactory, the machine in question may be accepted or prepared for shipment. However, if the study reveals unsatisfactory results, be prepared to explain the discrepancies or unacceptable portions of the study in an understandable and convincing manner. This, perhaps, is the most critical point in the entire study. You are, in effect, saying the machine is "no good", and the machine builder will be ready to defend his position that they build high quality equipment. It is probable that his first defense will be an attack on the accuracy of the probability study. Due to lack of knowledge of S.Q.C. principles, the machine builder will then question the reliability of every procedure that was used, including the final analysis. The gages will be questioned, and also the condition of the parts that were machined and used for the study. Usually a suggestion will be made to use selected parts. Be careful not to fall into this trap. Selected parts may show what the machine is capable of doing under ideal conditions, but it is important to keep in mind that pieces coming to the machine in the plant

56

will not be selected, but will be normally produced pieces with variability as great as we can possibly stand from an economical point of view.

3. The foregoing steps were suggestions in making a capability study in the machine builder's plant. However, the study is not complete until the machine is released to production. The Quality Control man will need to work closely with the factory representative and plant personnel to see that the machine repeats its qualification test. Here again, care should be exercised to see that everyone concerned understands each step to be taken in making a good capability study.

It might be well to mention here that many additional benefits are obtained as a result of making good capability studies. A partial list would include:

1. Increase in number of good pieces produced
2. Reduced skill requirements
3. Higher morale
4. Increased pride in workmanship
5. Reduced scrap and rework
6. Reduced cost

Now, to show you how we went about making a capability study, I would like to refer to charts which give a pictorial story of the study.

Figure 1-A shows the top view of the grinder with the relative position of the part (distributor shaft) with the grinding wheel and feed wheel. Three (3) dimensions must be controlled — the cam end, .3112 to .3115; the bearing, .3732 to .3735; and the body, .4896 to .490. In addition, taper, out of round, and finish must also be controlled.

Figure 1-B shows the profile bar which controls the profile of the grinding wheel.

Figure 2 shows how the first pieces produced measured. These were recorded in a tally form which immediately indicated that the dimensions were not at the same level. The .3112/.3115 dimension was running too high; the .3732/.3735 and .4896/.490 were running too low. It can be readily seen that if an adjustment is made to reduce the .3112/.3115 dimension, the other two dimensions will go still farther below the lower specification limit. On the other hand, if adjustments are made to increase the .3732/.3735 and .4896/.490, the .3112/.3115 dimensions will go farther above the upper specification limit. In order to line up the dimension levels with respect to each other, it was necessary to take the following corrective action.

1. Profile bar required lapping on the machine.
2. Resting blade had to be realigned.
3. Wheel grit was changed.
4. Extra hard diamond for dressing had to be used.
5. Two (2) passes instead of one were required to dress grinding wheel.
6. Feed wheel required adjustment.
7. Master cam which governs feed wheel contour had to be reground.
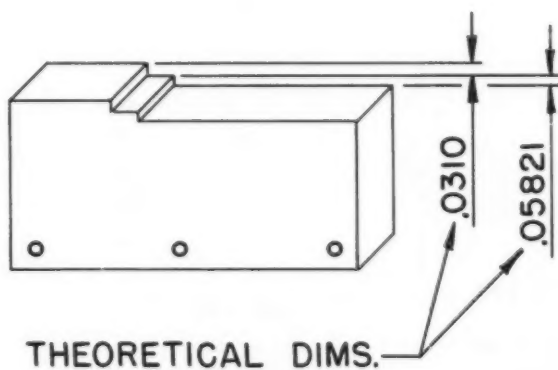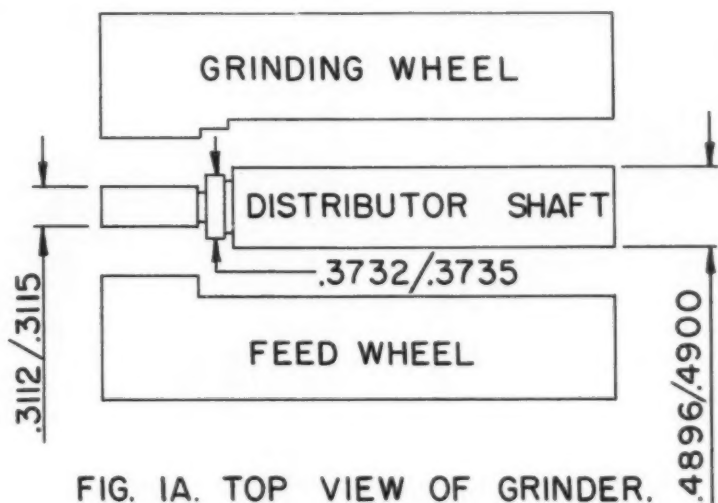8. Few other adjustments were required.

57

Following the preceding rework, the following study shows the results accomplished.

See Figure 3.

The last three charts, Figures 4, 5, and 6, show how the grinder is reacting today, two years later. As you can see, the operation is well in control.

In addition to the above capability study, an endurance run was made to determine the frequency of wheel dress requirements. Whereas it was an established practice to dress the wheel every 80 to 100 shafts, studies revealed that the wheel could grind approximately 450 shafts between wheel dressings.

We at Delco-Remy feel confident that the time required to make capability studies on new or reworked equipment is time well spent and adds immeasurably to the quality of our products.
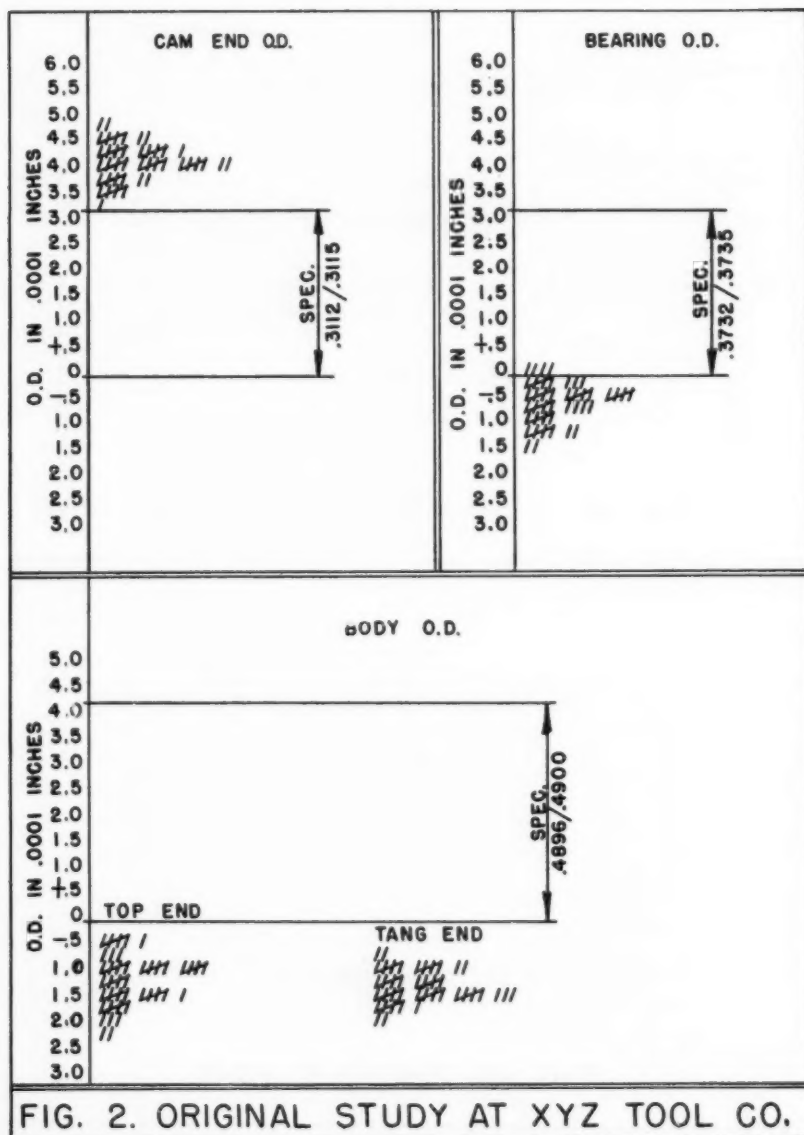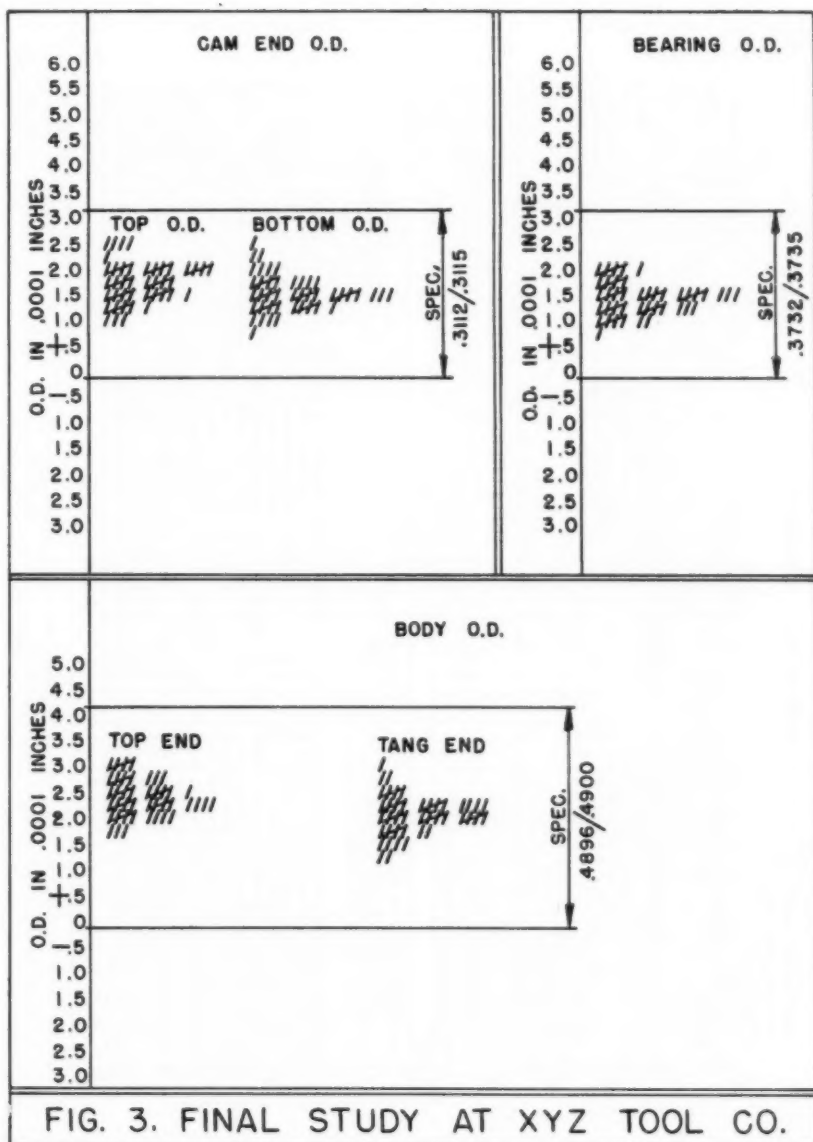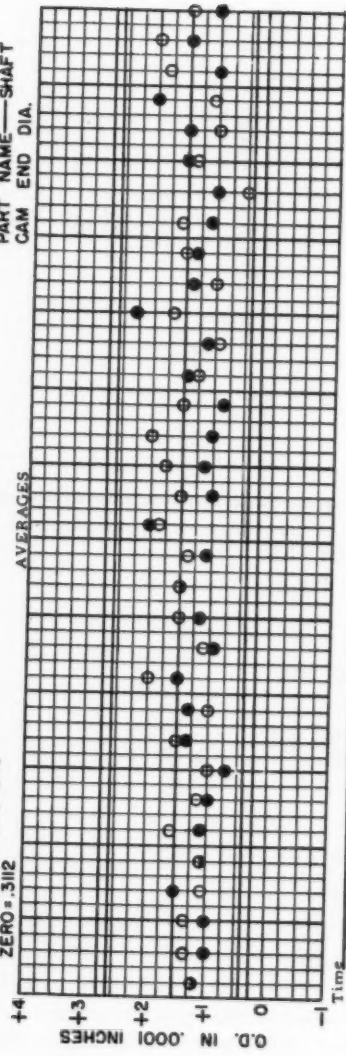
GRINDING WHEEL

DISTRIBUTOR SHAFT

.3732/.3735

FEED WHEEL

.3112/.3115

.4896/.4900

FIG. IA. TOP VIEW OF GRINDER.

.0310

.05821

THEORETICAL DIMS.

FIG. IB. PROFILE BAR.

FIG. 2. ORIGINAL STUDY AT XYZ TOOL CO.

FIG. 3. FINAL STUDY AT XYZ TOOL CO.

61

FIGURE 4.

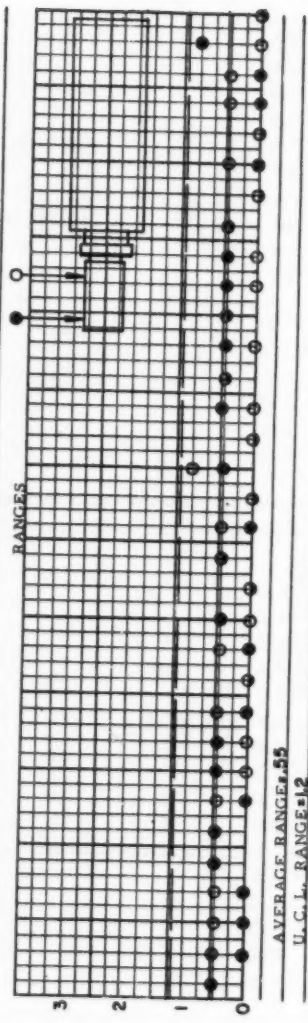62

CONTROL CHART

PART NO. 1923356
TOLERANCE=.3732-.3735
ZERO=.3732

SAMPLE SIZE=5
PART NAME—SHAFT
BEARING DIA.

AVERAGES

O.D. IN .0001 INCHES

+4 +3 +2 +1 0 -1

Time

GRAND AVERAGE
C. L. AVERAGES   UGL=2.54   LGL=.46
C. L. INDIVIDUALS

MODIFIED LIMITS

RANGES

AVERAGE RANGE=.65
U.C.L. RANGE=1.4
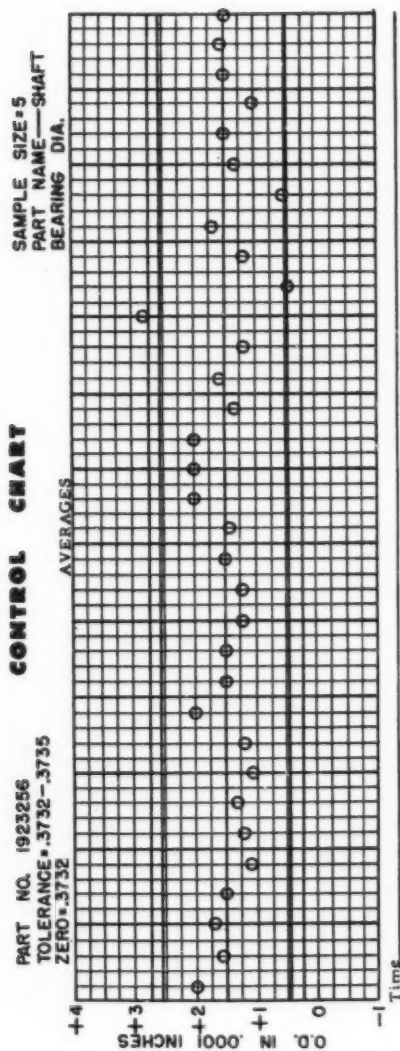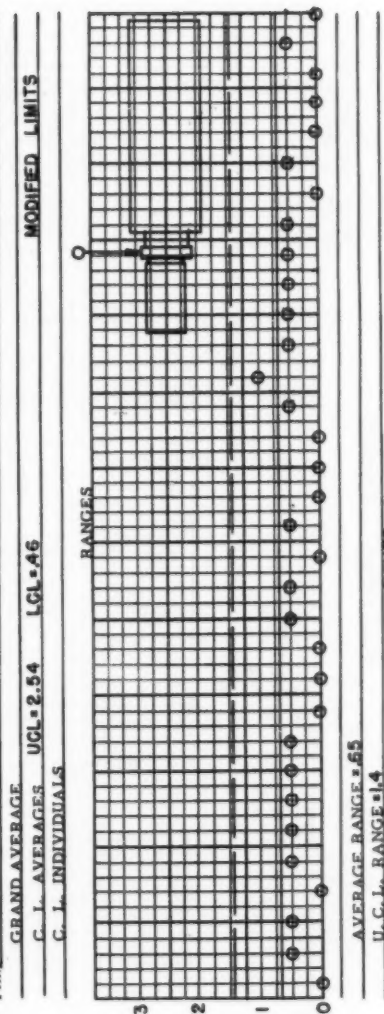
FIGURE . 5.

PART NO. 1923256
TOLERANCE=.4896–.4900
ZERO=.4896

CONTROL CHART

SAMPLE SIZE=5
PART NAME —SHAFT
BODY DIAMETER

AVERAGES

O.D. IN .0001 INCHES

+4
+3
+2
+1
0
-1 Time

GRAND AVERAGE
C. L. AVERAGES   UCL=3.37   LCL=.53   MODIFIED LIMITS
C. L. INDIVIDUALS

RANGES

3
2
1
0
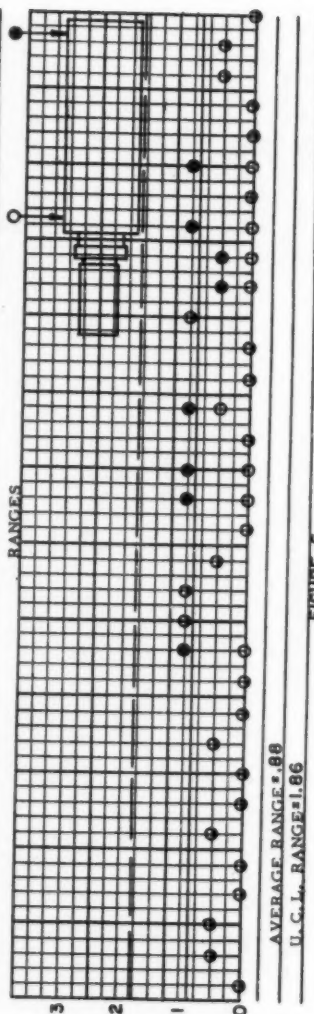
AVERAGE RANGE=.88
U. C. L. RANGE=1.86

FIGURE 6.

64

TRIAL RUNS FOR QUALITY INSURANCE

Harold J. Gulde
Minneapolis-Honeywell Regulator Company

Introduction

The developmental phase of a new product probably has more to do with later reliability of the unit than any other one thing. At Honeywell a new unit must pass through a design phase, a product planning phase and a trial run phase before it is ready for regular production.

Designs of new units are originated to satisfy a customer's specific needs or as a result of engineering research to find a better control. Two trial runs are made during this phase of the development. These are the building of engineering models and the testing of models in field application. Attention in this paper will be centered on later production trial runs rather than on trial runs during the design phase.

The product planning phase accomplishes the following things:

1. A preliminary classification of piece part characteristics is made,
2. Processes are assigned or developed,
3. Tools are ordered,
4. Gages are ordered,
5. Cost estimates are made,
6. Method layouts are written,
7. A trial run inspection procedure is written.

Piece Part Trial Runs

During the piece part trial run it is determined that piece part tools and processes are capable of meeting print tolerance. These trial runs may involve a machine capability study on the more critical dimensions. The critical dimensions were spelled out in the classification of part characteristics during the product planning phase. The number of parts checked from the trial run will vary from a very few on Punch Press operations to many on critical machining operations where the tool may wear rapidly.

The final piece part inspection procedure is completed at the time of the trial run of the piece part; it is based on the preliminary classification and on the trial run. The procedure is a list of characteristics to be checked, an AQL for each and the gages for each. The AQL keys the sampling plan to be used.

The First Trial Run of Units

Trial runs may be made to establish any one of the following:

1. That parts, all of which are to print, will assemble into units which meet specifications,
2. That parts, from production runs of all piece parts, will assemble with normal assembly practice into units which meet specification,
3. That future runs of production parts will assemble with normal assembly practice into units which will meet specification.

It must be decided which basis will be used for the trial run. Usually number three above is preferred although occasionally during this first trial run of the unit, sorted parts may be used per number one above.

The first trial run of the units is under the supervision of Process Engineering. It is here that the ability to build the unit is really tested. The problem is, will the unit when produced under production conditions give reliable results, both immediately when received in the field and also over a predetermined life. The process capabilities are measured, gages are proved out, and the testing for repeatability is completed at this time.

A fairly large group of units is processed (about 50-1000) under the watchful eyes of Honeywell engineers. Any irregularity is traced down before the unit leaves this stage. A preliminary inspection procedure on the unit is completed at this time. This procedure lists each characteristic and classifies it as Critical, Major, or Minor. This procedure will form a basis of the final inspection procedure in later runs of the unit.

Units from this run are also put on life and shipping tests. Life tests are accelerated tests which show how well the unit will stand up under normal and adverse conditions.

An example of the work done during the trial run is the repeatability - fixture - process variability study. Fifty units are numbered and read through on a fixture by an operator. On repeated readings of the fifty units, all previous readings are not available. Table I shows the results recorded from such a study. Table II shows the information to be obtained from means. Table III shows the variations measured by the example.

Table I
Repeatability - Fixture - Process Variability Study

Unit Number	Fixture 1			Fixture 2		
	Reading 1	Reading 2	Reading 3	Reading 4	Reading 5	Reading 6
1	9.5	10.1	9.6	10.1	12.7	13.0
2	9.2	9.5	10.4	11.2	12.5	12.7
3	7.1	8.2	8.6	9.1	10.1	10.3
4	8.6	9.1	9.1	8.6	9.9	10.0
5	10.3	12.3	11.5	11.4	12.5	12.0
6	7.2	7.4	8.5	9.5	10.6	10.2
7	10.9	12.1	12.8	11.0	11.8	11.5
8	9.0	10.5	10.2	10.0	11.9	11.0
9	9.2	9.0	9.3	10.8	10.7	11.3
10	8.1	8.6	8.5	10.7	10.9	10.8
-	-	-	-	-	-	-
-	-	-	-	-	-	-
-	-	-	-	-	-	-
$\bar{X}$	8.9	9.7	9.9	10.2	11.4	11.3

66

Table II
Mean Results from
Repeatability - Fixture - Process Variability Study

Information Obtained from Means	Means to Use
1. Over-travel in the first reading.	1. Reading 1 vs. Reading 2.
2. Drift in the unit.	2. Reading 2 vs. Reading 3. Reading 5 vs. Reading 6.
3. Comparison of fixture mean reading.	3. Readings 1, 2, 3 vs. Readings 4, 5, 6.
4. The process average.	4. Reading 2 or Reading 5.

Table III
Variation Results from
Repeatability - Fixture - Process Variability Study

Information Obtained from Variation	Variation to Use
1. Repeatability of Fixture 1.	1. Paired differences Reading 2 and Reading 3.
2. Repeatability of Fixture 2.	2. Paired differences Reading 5 and Reading 6.
3. Process variation.	3. Variation within the 50 unit readings of Reading 2 or the 50 unit readings of Reading 5.
4. Severity limit necessary between Operating and Inspection.	4. Paired differences between Reading 3 and Reading 5.

Production Acceptance Run - The Final Trial Run

The next hurdle for the unit is the Production Acceptance run. In this phase the unit is built by production people under production supervision. All assembly tools and gages must be performing satisfactorily at this time. This run is usually fifty units, but on important developments it may be as large as 1000 units.

The process variation on each assembly characteristic is checked again. If the distribution is normal it must be comfortably inside of specification. It is not enough for the six σ spread to be just inside of specification. This is particularly true with a characteristic that is non-adjustable such as wattage. In such case, a new set of parts can easily shift the distribution. An adjustable characteristic however, such as calibration can usually have the six σ limits nearer specifications. Also, this type of characteristic is easier to salvage once a rejection has occurred.

The inspection procedure on the unit is completed at this time; each characteristic classification is approved. The sampling plans are finalized and the inspection stations are set up.

A sample from the run is sent to Engineering. Floor results are correlated with results on Engineering Test fixtures. There are also

many additional design tests run at this time over and above those run on the floor.

Summary

A unit developed by Honeywell and in production must successfully pass the following trial runs:

1. Engineering models,
2. Test of Engineering models,
3. Field models tested in actual use,
4. Trial run of piece part,
5. Trial run of unit,
6. Life Test,
7. Shipping Test,
8. Production Acceptance run,
9. Engineering tests of final unit.

DISTRIBUTION REQUIREMENT SPECIFICATION

B. J. Kinsburg
Bell Telephone Laboratories

1. Introduction

The distribution requirement specification which is presented in this paper was worked out to meet a specific need of the L3 Coaxial Telephone System. The L3 System is a long distance carrier system designed to transmit either 1,860 telephone channels, or 600 telephone channels plus one television channel. The band width is approximately 8 mc. The amplifiers are spaced along the route at 4 mile intervals. In a 4,000-mile route there will be about 1,200 amplifiers in tandem including those at the terminals.

These repeaters or amplifiers supply a total of about 54,000 db of gain to make up for 46,000 db loss in the coaxial cable and 8,000 db loss in the associated equalizers at intermediate offices and terminals. The required match between the total amplifier gain and the cable and equalizer loss must be held to better than 1 db for a satisfactory system performance. The practical realization of such performance was made possible only through the application of the distribution requirements on the components of the system; and in particular through a close control of process averages.

To place the distribution requirements in a proper perspective two important steps in the design of the L3 System should be mentioned.

Deviations Studies

The ·ircuit designs were analyzed to determine the sensitivity of the circuit to variations in the component elements. In case of the L3 System the main characteristic of interest was the deviation of amplifier gain from its central or design value. The need for a quality control scheme to limit excursions of the long term process average from the central value was established early in the design stage and a Committee was appointed to work out a solution. The work of the Committee eventually resulted in the distribution requirement specification.

Selection of Elements - Equivalent Limits

Let us assume that each element in a circuit is allotted an equal share in its contribution to the variability pattern of the system. Using this assumption and the results of the deviation study it is possible to calculate the required tolerances on each component element. In actual practice what was desired is the most economical overall solution to the problem of distribution control. The tolerances so obtained were therefore modified to permit some elements a greater and some a lesser share in contributing to overall distribution depending on the inherent variability of each manufactured component element. These tolerance requirements on component elements were arrived at in consultation with the element designers. If available types of elements were not capable of providing the desired performance, new types of components had to be developed. Of interest to the system designer were not only manufacturing tolerances, but also items such as reliability and variability with age and temperature. Thus arriving at a satis-

factory solution required a large amount of give and take between the circuit and the component elements designers.

2. The Basic Decisions

The committee which worked out the distribution requirement specification consisted of Bell Laboratories and Western Electric Engineers with Mr. H. F. Dodge as chairman.*

At the outset certain decisions were made which broadly defined the content of the specification. These decisions were:

(1) To limit, with a fair degree of reliability, the systematic variations in the process average of the specified characteristic of a product from a design center or a nominal value (N) within a band defined by $\pm 1/3\ \sigma$ where σ is determined by the basic capability of the process.

Thus: PA $lim = N \pm 1/3\ \sigma$

(2) To accept all product within $\pm 3\frac{1}{3}\ \sigma$ from the nominal value. Thus the limit on the individual values which will be called the A limit becomes:

A limit $= N \pm 3\frac{1}{3}\ \sigma = N \pm A$

Incidentally, this second decision leads to a simple numeric for the limit on the process average since now we can express

PA $lim = N \pm \frac{1}{3}\ \sigma = N \pm 0.1A$

The above two decisions are illustrated by Fig. 1. As shown on Fig. 1 our discussion will be confined to the case of characteristics having substantially normal distributions and having both minimum and maximum specified limits. In the overall plan, however, provisions have been made for characteristics having skew distributions or having a single specified limit, maximum or minimum. Fig. 1 shows two limiting acceptable distributions, normal in shape, both having a standard deviation σ'' equal to the standard deviation of the process, one having an average, X'' located $1/3\ \sigma''$ below nominal and the other having an average $1/3\ \sigma''$ above nominal. The allowance of $\pm 1/3\sigma$ is, of course, arbitrary and represents an estimate of relative importance of the systematic to random changes in the process average.

(3) The third decision made by the Committee was broader in character than the two preceding and, therefore, more difficult to define. Briefly stated it was to provide sufficient flexibility so that the producer would not be unnecessarily limited in his choice of the manufacturing process. Also it is necessary to furnish the producer with a method of operation while the process is out of control.

- -

In order to implement the distribution requirements it is necessary to provide procedures which define: a) The character and quantity of evidence needed regarding the collective quality of a product as it is made day by day. Of necessity such procedures are of the nature of inspection practices. b) The criteria for judging when such product may be conforming to the intent of the specification, and c) The treatment or disposition of product units when these criteria are not met.

To implement the above decision, several procedures or methods have been prepared to meet the different conditions that may be encountered in production. We will discuss, with particular emphasis on the intent, four of these methods:

a. control chart method
b. batch method
c. three-cell method
d. records method

(4) In some cases it would be desirable from the design point of view to apply distribution requirements to more than one characteristic of the product. For instance, it is desirable to control the process average of the gain-frequency characteristic of the amplifier at several frequencies. Another example is in the case of combined elements in which a coil is wound with resistance wire so that the resulting element has both resistance and inductance.

It is easy to see that the administration of distribution requirements in such cases may be quite cumbersome when one of the characteristics is out of control. The decision made in this case was to select one characteristic for application of distribution requirements and apply only go-no-go limits coupled with the records method to the remaining characteristics.

3. Control Chart Method

Outline of the Method

This method was designed to apply to an essentially continuous production process in which the sequence of individual units is not lost. To quote the specification it "is intended for application where production comprises a reasonably steady succession of individual units or small group of units from a common source; so that individual units, as produced, may be kept in the order of their production and control techniques applied to test results".

The control charts specified are an adaptation of the Shewhart control charts for sample averages, $\bar{X}$, and sample ranges, R, for "control with respect to a given standard" (see Part 3 of the ASTM Manual on Quality Control January 1951). Two modifications in the control chart for $\bar{X}$ have been introduced: a) A central band rather than a central line has been provided, and b) a run requirement has been imposed on seven successive sample averages. The sample size has been standardized at 5.

The limiting distributions for sample averages are illustrated on Fig. 2. Referring to Fig. 2 the desired process average limits (PA limits) are the averages of the two acceptable distributions of averages of 5 units.

71

PA limits = $N \pm .1A$

The limits for individual averages of samples as limited are derived as follows:

$$A5 \text{ limits} = N \pm \left(1/3\, \sigma'' + 3\, \frac{\sigma''}{\sqrt{5}} \right)$$

$$= N \pm (.1A + .4A)$$

$$= N \pm .5A$$

The limit for ranges (R5) is the customary upper 3-sigma control limit for ranges to be met by the range sample of 5

$$R5 \text{ limit} = 4.918\, \sigma''$$

$$= 4.918\, \frac{3}{10}\, A$$

$$= 1.48A \text{ max.}$$

At the outset, and whenever eligibility to use the control chart method is lost, a 100 per cent inspection rate is required. To establish eligibility for use of the control chart method, at least 7 consecutive samples must satisfy the following criterion.

Criterion I - Establishing Eligibility

    a) The averages $\overline{X}$, all fall within the A5 limits.

    b) The ranges, R, all fall within R5 limits.

    c) Seven consecutive averages, $\overline{X}$, are not all outside the same PA limit (not all above the upper PA limit or all below the lower PA limit).

Eligibility for use of the control chart method is maintained so long as each current sample satisfies Criterion II.

Criterion II - Maintaining Eligibility

    a) The average, $\overline{X}$, either

        (1) falls within the A5 limits; or
        (2) falls outside the A5 limits, but at the same time all of the 6 preceding consecutive averages fall within for A5 limits;

    b) The range, R, either

        (1) falls within the R5 limit; or
        (2) falls outside the R5 limit, but at the same time all of the 6 preceding consecutive ranges fall within the R5 limits;

    c) Seven consecutive averages, $\overline{X}$, (for the current sample and the 6 preceding samples) do not all fall outside the same PA limit.

d) No major change is made in raw material, machine set-up or personnel, which, in the judgement of the functionally responsible personnel, may have a significant effect upon the quality of the product.

Loss of Eligibility

When the current sample fails to satisfy Criterion II the three-cell method applies to the product represented by the current sample, and to subsequent production until eligibility is re-established by satisfying the conditions of Criterion I.

Operating Characteristics of the Control Chart Method

The operating characteristics of the control chart method are shown on Fig. 3. In computing these curves the statistical models used to evaluate the expected performance were limited to normal distributions. The ordinate is the probability of accepting the current sample and the abscissa is the displacement of the process average, $\bar{X}$, from the nominal value, N.

Fig. 3a shows the OC curves for Criterions I and II when the standard deviation of the process $\sigma' = 0.3A$. The Criterion I is seen, as would be expected, to be more stringent than Criterion II.

Fig. 3b shows what would happen if the non-parametric requirements on the 7 consecutive points were omitted. These curves are labelled "I, II, less C". It is seen that without this requirement Criterion I and particularly Criterion II would be very ineffective in controlling the departures of the process average from the desired value (N).

'Fig. 3c shows the operating characteristics for Criterion I when $\sigma' \neq 0.3A$. This would happen if the original specified limits were improperly set or it the process dispersion changes. Similar curves for Criterion II are plotted on Fig. 3d. As would be expected, Criterion II is not as effective in constraining departures of the process average from N.

4. Batch Method

Outline of the Method

This method is intended for application where production consists of intermittent batches of 50 or more units all of which have been made under the same essential conditions with respect to materials, parts, workmanship and processing. The assumption is that the time sequence of production of individual units even if it could be traced is not significant and the only order which has a meaning is the sequence of successive batches. It is also assumed, however, that there is a basic continuity in raw materials, parts and workmanship, which is insured by appropriate quality control method.

In this method a sample of 50 units is selected at random from each batch. The average, $\bar{X}$, and the standard deviation, σ, of the sample are computed and the acceptability of the batch is determined by Criterion III.

Criterion III - Batch Acceptability

    a) The average $\bar{X}$ meets the A50 limits where

$$\text{A50 limits} = N \pm (1/3\ \sigma^m + 3\sigma^m/\sqrt{50})$$

$$= N \pm 0.23A$$

    b) The standard deviation, σ, either

        (1) falls within S50 limit where S50 limit

            = .41A, max, or

        (2) fails to meet the S50 limit, but at the same time all of
the 6 preceding consecutive standard deviations meet the
S50 limit.

The basis for A50 limits is shown on Fig. 4.

Operating Characteristics of the Batch Method

The operating characteristics for the batch method when $\sigma' = 0.3A$
is shown on Fig. 6a. For comparison purposes this figure also shows the
corresponding curve for Criterion II of the control chart method.
Figure 6b shows operating characteristics for several values of σ' rela-
tive to A. As would be expected, the batch method, due primarily to the
use of a relatively large sample, is somewhat more efficient in con-
trolling the process average than the control chart method.

5. Three-Cell Method

Description

Even the best behaved process is capable of going out of control
and staying out of control for a considerable period until the assign-
able cause or causes are found and corrected. If in the interim the
shipments of the finished products stop, the effect on the interlocking
schedules may be disastrous. Also what can the harassed manufacturing
organization do with the uncontrolled product already on the shelves?

The solution of this problem is the three-cell method and is due to
the genius of H. F. Dodge. It provides a procedure for continuity of
manufacture when the product is out of control. This is a procedure
which permits shipment of some of the product and at the same time in-
sures the consumer that the shipped product will have a satisfactory
distribution.

The three-cell method may be used at the discretion of the manu-
facturing organization in place of either the control chart or batch
method. The use of this method is of course mandatory when the product
fails to meet the criteria of the other methods.

The basis for the three-cell method is shown on Fig. 5. Each unit
of the product is tested and conforming units are classified in one of
three cells: lower, middle or upper. The product is shipped in groups
of 5 units, each group consisting either of 1 unit from the lower cell,
3 units from the middle cell and 1 unit from the upper cell or all five

units from the middle cell. In this way each group of five units forms a small distribution of its own.

In addition to sorting the conforming units in three cells and requiring shipment only of a selected distribution this method also requires maintenance of control charts on average ($\overline{X}$) and range (R) using random samples of 5 units selected in the order of production prior to classification. In this way a record is kept of the process showing when corrective action should be taken.

Operating Characteristics of the 3-Cell Method

The corrective effect of the 3-Cell method on the process average when $\sigma' = 0.3A$ is shown on Fig. 7. In this figure the abscissa is the deviation of the process average (x') from the nominal (N) and the ordinate the deviation of the average of the packaged product. The solid line shows the average long term corrective effect produced by packaging. The other two curves show the corrective action if either only 0-5-0 packaging (dashed line) or only 1-3-1 packaging (dot-dash line) were employed. The efficiency of the method is seen to be a factor of approximately 3. This efficiency is obtained at a price as seen on Fig. 9. In this figure the abscissa is, as in the previous figure, the deviation of the process average. The left hand scale applying to the upper (solid) line is per cent of product packaged. The right hand scale applying to the lower (dotted) line is the average of the packaged product. For instance, when the process is centered 0.2A away from the nominal value, only 70% of the product can be packaged. At the same point the average of the shipped product is about .07A away from the nominal, thus confirming the 3 to 1 efficiency factor. Of course, if the process exhibits a shift in the opposite direction it may be possible to ship the unpackaged units. This incidently furnishes an incentive for corrective action.

One of the pitfalls of the 3-Cell method is seen on Fig. 8, where the effect of distributions in which $\sigma' < 0.3A$ is shown. This may occur if the original limits are too lenient or if a change in the process takes place resulting in a tighter distribution. On Fig. 8 the abscissa is the deviation of the process average from the nominal and the ordinate the deviation of the packaged product. For instance when $A = 8\sigma$ the 3-Cell method performs essentially no corrective action. This is an additional reason why maintenance of conventional control charts to supplement the sorting operation is required.

6. Records Method

The records method primarily involves maintenance of control charts of the product. This method is intended for application when it is not practicable to apply the other methods which were described above but at the same time some control is desirable. The records method is usually specified when there are several characteristics of the product in which it would be advantageous from the system point of view to minimize deviations of the process average from the nominal value. In such cases the most important of these characteristics is selected for application of other methods and the records method specified for the remaining characteristics. An example are the combined elements in which a coil is wound with a resistance wire on a ceramic spool. The resulting unit has to meet both inductance and resistance limits. In this case the inductance was selected for application of the control chart method (or the three cell method if preferred by the manufacturer) and the resistance was administered under the records method.

75

In effect this is a "pretty please" method in which the manufacturer is asked to control the product but the action taken to achieve control is left to his initiative. At the same time it has two important advantages;

1) The customer and the producer have the evidence of control charts.

2) The producer knows from the manufacturing specification that control charts are <u>required</u> and therefore includes them in his layout of the process and in the pricing structure.

7. Accuracy of Measurements

Any framework of distribution requirements would be incomplete without taking into account the accuracy of measurements. It is convenient to separate the accuracy of measurements into two major components: precision and **bias**. By definition precision contains all random contributions and may be assumed (unless contrary evidence exists) to be normal in character. The bias contains all systematic components.

The effect of precision is illustrated on Fig. 10. As seen from this figure the effect of the normally distributed error of measurements is to produce an apparent distribution of the measured values which is greater than the distribution of the true values. Let us designate the measured distribution by q and the distribution of true values by r and the distribution of measuring error by s.

The relationship between q, r, and s may be readily calculated from

$q = \sqrt{r^2 + s^2}$. The calculated rations of q/r for given rations of s/r are listed in Table I below.

Table I

Effect of Precision

Ratio s/r	Ratio q/r
1	1.414
.5	1.118
.2	1.020
.1	1.005

In the distribution requirements specification the overall instrument accuracy not poorer than 1/6A was specified with allowance of 1/10A for precision. This in effect allows 1/15A for bias.

In the application of the distribution requirements to specific components the battle of accuracies had to be won first. As a rule the process capability studies had to be made and in many cases new measuring equipment had to be developed before adequate precision was achieved. Improvement of measuring equipment or techniques in addition to achieving required degree of precision also tends to reduce bias to tolerable values. In some cases, however, comparison standards had to be introduced. Use of comparison standards is an art in itself and is outside the scope of this paper.

8. Application

The application of the distribution requirements to the L3 System was a step by step process. These steps are outlined in brief below in order to furnish some insight as to how distribution requirements fit into development of a complex transmission system.

Component Elements

The use of the distribution requirements was specified for all critical component elements in the L3 amplifier. Some of these were specially designed for use in these amplifiers because the available elements did not meet the system objectives based on early deviation studies. Each instance of realization of these special designs is an interesting quality control story in itself. Some of the quality control aspects of this work in connection with the vacuum tubes were published in AIEE transactions.* An early summary of application of distribution requirements to manufacture of other component elements is also available.** As in all development work most of the effort remains untold though the results have been incorporated into the general pattern of the Bell System.

Field Trial

In this stage a field installation was equipped with amplifiers made in the factory on a preliminary production line utilizing components which met distribution requirements. The resulting system was subjected to comprehensive tests. Any significant deviations from design center values were examined to ascertain assignable causes. All cases of failure were analyzed, and necessary remedial action was taken.

Production

Initially the distribution requirements were accepted by the Western Electric Company for the component elements only. From the very beginning, however, the overall amplifier gain characteristics were watched for conformance with the distribution requirements. The control charts were being plotted and corrective action when required was initiated by production engineers. After about two years of production experience the distribution requirements were incorporated into the manufacturing specifications. The control chart method was applied to the amplifier gain at the top of the transmission band, and the records method was applied at two other frequencies, one in the middle and one at the bottom. A typical control chart of the overall amplifier gain at 8.32 megacycles is shown on Fig. 11. The excellent control displayed in Fig. 11 is not an accident but the result of long and painstaking effort.
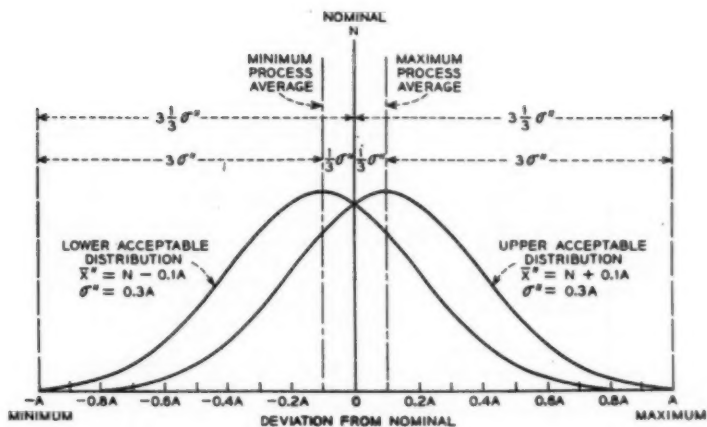
---

* Application of Statistical Techniques to Electron Tubes for Use on a 4000 Mile Transmission System - by W. Van Haste AIEE Transactions Part I No. 23, pp 50-54, March 1956.

** The L3 Coaxial System - Application of Quality Control Requirements in the Manufacture of Components - by R. F. Garrett, T. L. Tuffnell and R. A. Waddell.

The Bell System Technical Journal, July 1953, pp. 969-1005.

BASIS OF SPECIFIED LIMITS
FIG. I
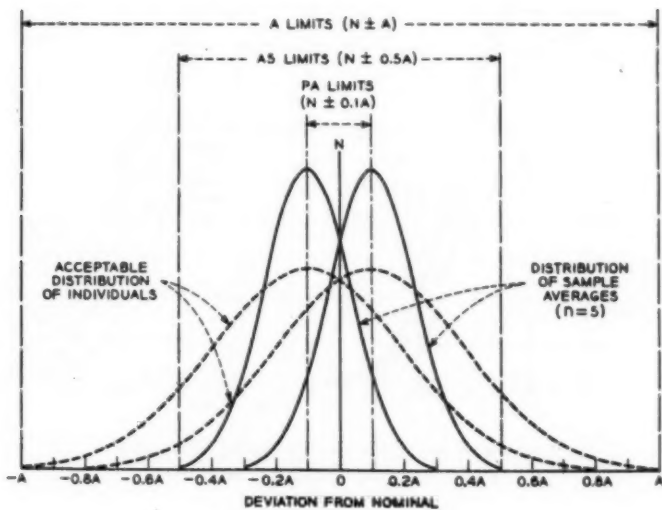
BASIS OF A5 LIMITS AND PA LIMITS
FOR SAMPLE AVERAGE
FIG. 2

79
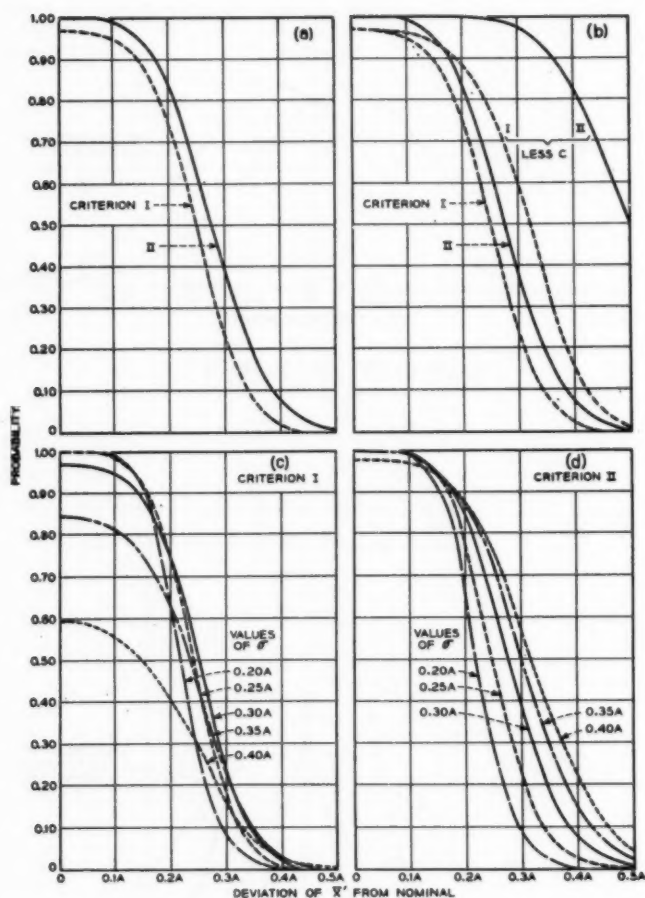
OPERATING CHARACTERISTIC CURVES
FOR CONTROL CHART METHOD
FIG. 3
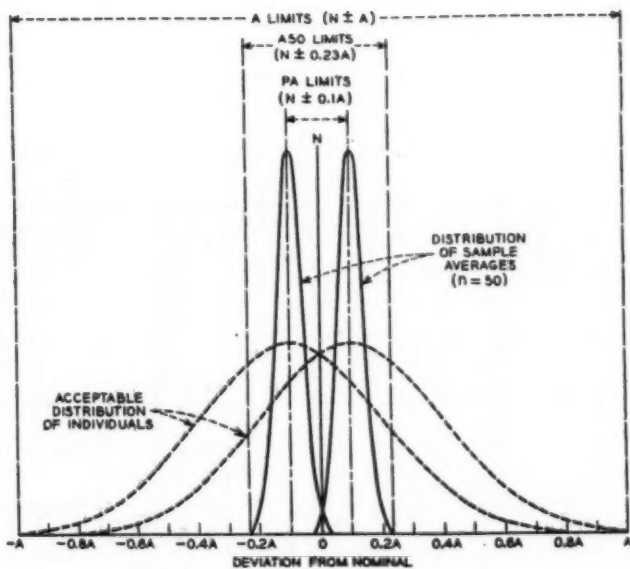
BASIS OF A50 LIMITS FOR SAMPLE AVERAGES
BATCH METHOD
FIG. 4
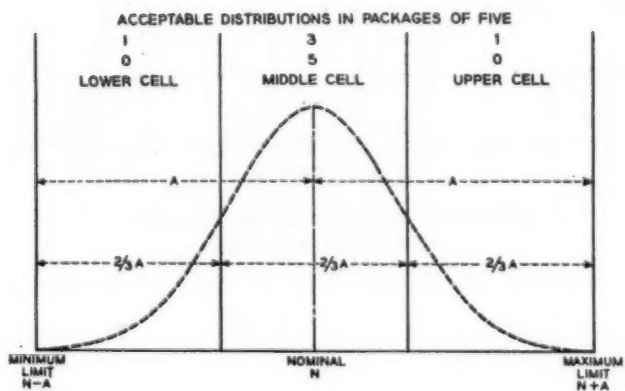
ACCEPTABLE DISTRIBUTION OF UNITS IN PACKAGES
OF 5, THREE-CELL METHOD
FIG. 5

81
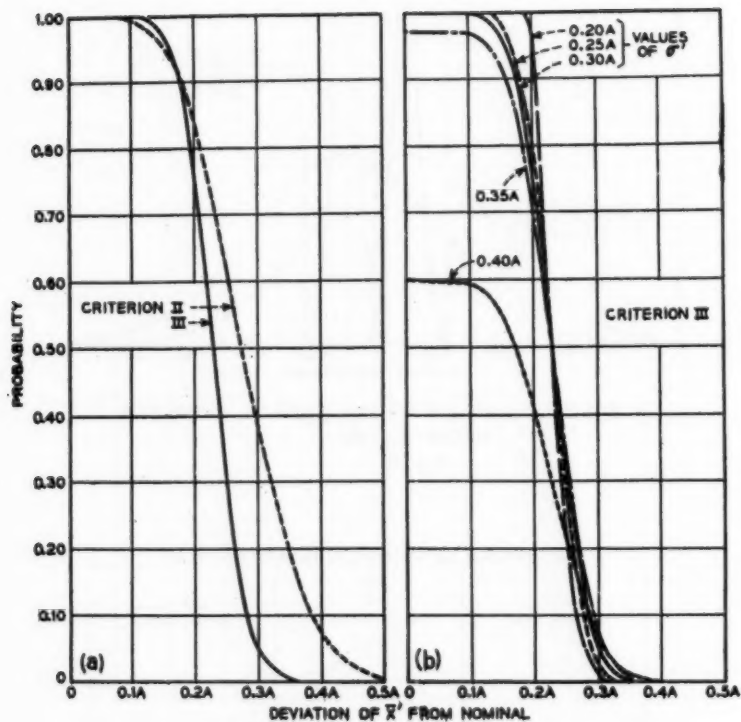
OPERATING CHARACTERISTIC CURVES
FOR BATCH METHOD
FIG. 6
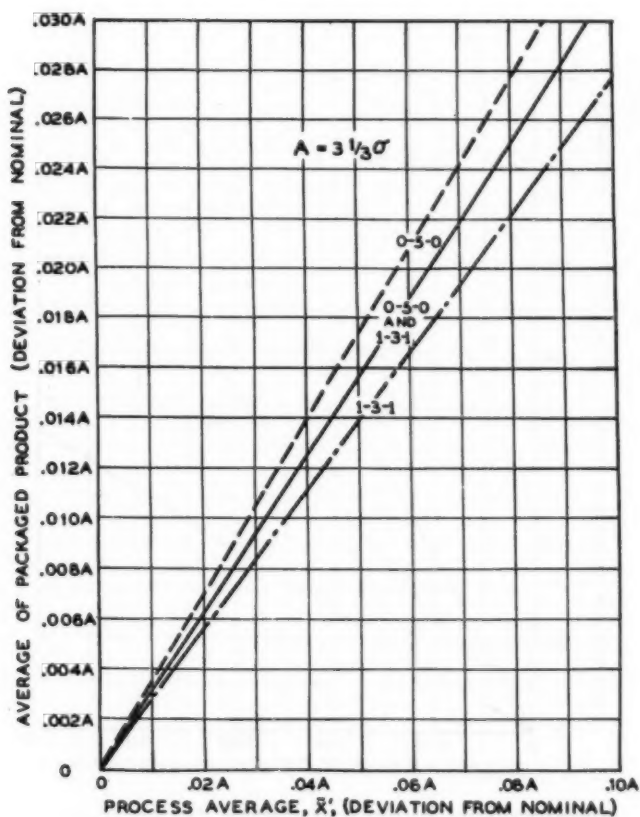
CORRECTIVE EFFECT OF THE THREE-CELL METHOD

FIG. 7
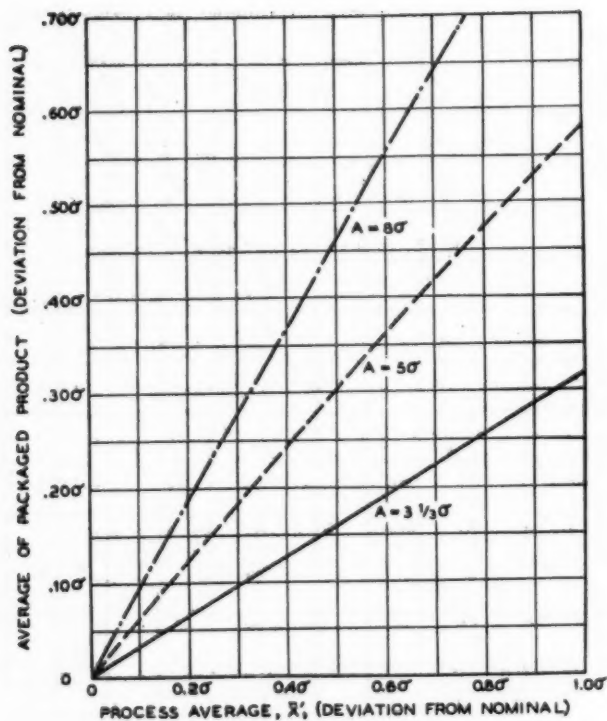
EXPECTANCY CURVES, AVERAGE OF PRODUCT
PACKAGED BY THREE-CELL METHOD
FIG. 8

84
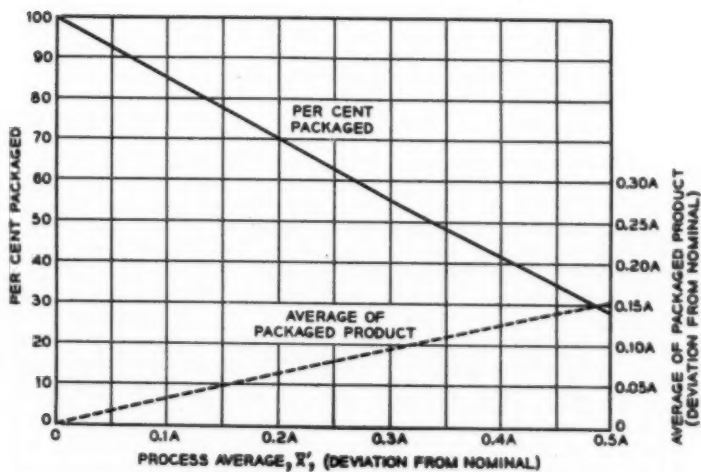
EXPECTANCY CURVES FOR THE THREE—CELL METHOD
FIG. 9
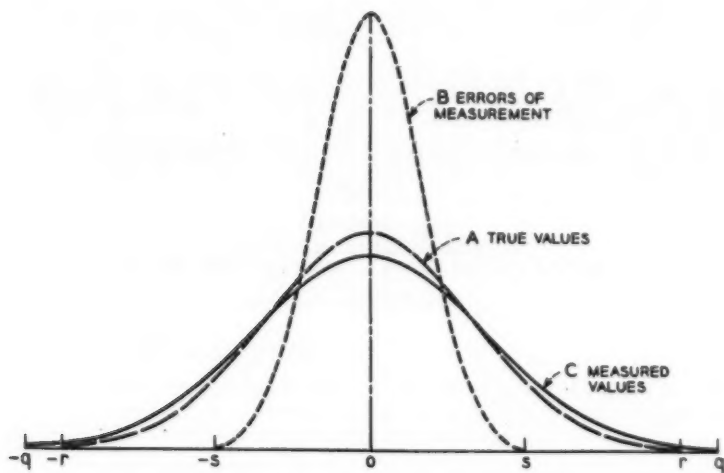
EFFECT OF MEASURING ERRORS
FIG. 10
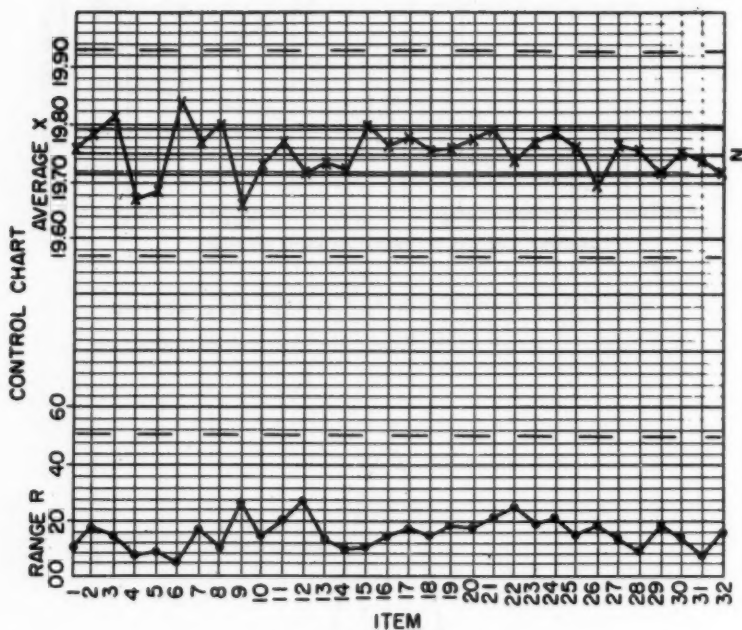
LINE AMPLIFIER
TRANSMISSION AT 8.320 MC
SPECIFICATION LIMITS: __ MIN. 19.39 MAX. 20.11

FIG. 11

OPTIMIZING A LIFE TESTING PROGRAM

Leonard G. Johnson
General Motors Corporation

I. Introduction

In recent years the science of operations research has been applied to a variety of economic situations. As a rule, in economic problems we are interested in either minimizing a cost or maximizing a profit. The operations research method of arriving at the maxima or minima of an economy is to set up a mathematical model of the system, and then proceed to analyze the mathematical equations for maxima and minima. Amongst the multitude of industrial problems lending themselves to this type of approach, we can find the very interesting and important question of how best to design a fatigue testing program for a given type of specimen, be it an automotive chassis part, a ball bearing, a railroad rail, or even a clock spring. Many other life testing programs fall into the same general category, including the testing of light bulbs, endurance tests on rubber tires, and corrosion and wear life tests. The design of a fatigue testing program involves two basic economic considerations, namely, time and money. Actually, the value of the time element can be expressed in dollars and cents, and, therefore, we should not consider time to be independent of money. However, since it is usually desirable to know how much time a testing program will consume, regardless of cost, it is advisable to analyze time and cost separately.

II. Assumptions

1. The fatigue life distribution of the item under consideration is a Weibull distribution.

2. The desired coefficient of variation in the mean life has been specified.

3. A fixed number of specimens are run simultaneously (with replacements as failures occur) until the test is completed. (The fixed number involved here shall be denoted by n).

4. The manufacturing cost per specimen is known. (Denote this cost by C_s).

5. The waiting cost per hour is known. (Denote this cost by C_w).

III. Questions to be Answered

1. How many specimens must be failed in order to give the desired coefficient of variation in the mean life? (The answer to this question will be a value known as r).

2. What is the median waiting time for a given value of $\left(\frac{n}{r}\right)$? (Denote this answer by $\left(\frac{n}{r}\right)\overline{|}_{.5}$).

3. What is the total cost of the testing program for a given $\left(\frac{n}{r}\right)$ and $\left(\frac{C_w}{C_s}\right)$?

4. For a given $\left(\frac{c_w}{c_s}\right)$, what value of $\left(\frac{n}{r}\right)$ will minimize the total cost?

IV. Properties of the Weibull Distribution

Before we can answer any of the questions listed in III, we must derive certain mathematical properties of the Weibull distribution which will be used as tools in obtaining the required answers.

Let X = the fatigue life of one specimen (in hours)

Then, according to Weibull [1], the probability that any specimen will fail within X hours is F (X), where F (X) is given by the formula

$$F(X) = 1 - Exp\left[-\left(\frac{X}{\theta}\right)^b\right] \qquad \binom{b > 0}{\theta > 0} \qquad (1)$$

(e is the base of the natural system of logarithms)
(b and Θ are known as the parameters of the Weibull distribution)
(b is called the Weibull slope)
(Θ is called the characteristic life)

By transposition we can write (1) as follows:

$$1 - F(X) = Exp\left[-\left(\frac{X}{\theta}\right)^b\right] \qquad (2)$$

By inverting both sides of (2), we obtain

$$\frac{1}{1 - F(X)} = Exp\left[\left(\frac{X}{\theta}\right)^b\right] \qquad (3)$$

Taking the natural logarithm of both sides of (3) gives

$$\log_e \frac{1}{1 - F(X)} = \left(\frac{X}{\theta}\right)^b \qquad (4)$$

Taking the natural logarithm of both sides of (4) gives

$$\log_e \log_e \frac{1}{1 - F(X)} = b \log_e X - b \log_e \theta \qquad (5)$$

88

Weibull probability paper is constructed by using (5) as a basis. Along the abscissa scale we list the value of log x, i.e., the abscissa is a logarithmic scale. Along the ordinate scale we list the value of log log $\frac{1}{1-F}$. We call this ordinate scale a Weibull scale of the probability F. Hence, with x plotted on a logarithmic scale, and with F plotted on a Weibull scale, we obtain a straight line of slope b. This explains why the parameter b is known at the Weibull slope.

Three important properties of any probability distribution are:

1. Mean
2. Median
3. Standard deviation

These can be expressed in terms of the parameters b and θ as follows:

$$\text{MEAN} = M = \theta \cdot \Gamma\left(1 + \frac{1}{b}\right) \tag{6}$$

$$\text{MEDIAN} = L_{.5} = \theta \cdot \left(\log_e 2\right)^{\frac{1}{b}} \tag{7}$$

$$\text{STANDARD DEVIATION} = \sigma = \theta \cdot \sqrt{\Gamma\left(1 + 2/b\right) - \Gamma^2\left(1 + 1/b\right)} \tag{8}$$

The life for which the probability of failure is q we denote by L_q. It is obtained by solving (2) for x when F (x) is taken equal to q. Thus,

$$L_q = \theta \cdot \left(\log_e \frac{1}{1-q}\right)^{\frac{1}{b}} \tag{9}$$

The standard deviation for a mean as estimated from a sample of r failures has been shown [3] to be equal to the population standard deviation divided by the square root of r. Thus,

$$\sigma_{\text{Mean}} = \frac{\sigma}{\sqrt{r}} = \frac{\theta}{\sqrt{r}} \sqrt{\Gamma\left(1 + 2/b\right) - \Gamma^2\left(1 + 1/b\right)} = \frac{M}{\sqrt{r}} \sqrt{\frac{\Gamma\left(1 + 2/b\right)}{\Gamma^2\left(1 + 1/b\right)} - 1} \tag{10}$$

$$\frac{\sigma_{\text{Mean}}}{M} = \frac{1}{\sqrt{r}} \sqrt{\frac{\Gamma\left(1 + 2/b\right)}{\Gamma^2\left(1 + 1/b\right)} - 1} = K \tag{11}$$

The ratio $\frac{\sigma_{\text{Mean}}}{M}$ = k is known as the coefficient of variation of the mean.

It is a measure of the accuracy of the estimation of the mean life from a sample of r failures.

According to assumption 3 in II, n specimens are running simultaneously. Consequently, any failure that occurs represents the first failure in a collection of n items. The median life of the first failure in n new items from a common Weibull population is obtained by

89

raising (2) to the power n, and solving for x when $\left[1 - F(x)\right]^n = \frac{1}{2}$.

Thus, MEDIAN LIFE OF FIRST FAILURE IN n $= {\left(\frac{1}{n}\right)}X_{.5}$

$$= \theta \cdot \left(\frac{\log_e 2}{n}\right)^{\frac{1}{b}} \quad (12)$$

The Answer to Question Number 1

The number of specimens which must be failed in order that a given coefficient of variation in the mean life can be realized is obtained by solving (11) for r.

Thus,
$$r = \frac{\dfrac{\Gamma\left(1+ 2/b\right)}{\Gamma^2\left(1+1/b\right)} - 1}{K^2} \quad (13)$$

The analytical relation (13) is plotted in Chart 1 for Weibull slopes b = 1, 1.5, and 2.

The Answer to Question Number 2

The median waiting time required in order to fail r specimens when n are kept running simultaneously (with replacements of failures) is the median of the sum of r time periods, each period being the elapsed time between two successive failures.

Equation (1) tells us that x^b is exponentially distributed with a mean value ρ^b.

Let $x^b = y$

Then $F(y) = 1 - Exp\left(-y/\theta^b\right)$ $\quad$ (14)

Since y is exponentially distributed, it follows that aging does not affect the distribution of future incremental values of y (See reference [4]).

Let y_1 = the value of x^b up to the 1st failure,

$\quad y_2$ = increment in x^b from the 1st to the 2nd failure,

$\quad y_3$ = increment in x^b from the 2nd to the 3rd failure,
.
.
.
$\quad y_r$ = increment in x^b from the (r-1)th to the rth failure.

The median value of each y_i (i=1,2, ..., r) is given by the b$\underline{th}$ power of (12), i.e.,

$${\left(\frac{1}{n}\right)}y_{.5} = \theta^b\left(\frac{\log_e 2}{n}\right) \quad (15)$$

Hence, the sum of the median values of $y_1, y_2, \ldots, y_r$ is r times (15), i.e.,

$$S = r \, \theta^b \left(\frac{\log_e 2}{n} \right) \tag{16}$$

The quantity S has dimension $(TIME)^b$. In order to reduce its dimension to that of TIME we take the bth root of (16). Thus,

$$[\text{SUM OF (MEDIAN INCREMENTS)}^b]^{1/b} = S^{1/b} = r^{1/b} \, \theta \left(\frac{\log_e 2}{n} \right)^{1/b} \tag{17}$$

In order to reduce (17) to the MEDIAN OF THE SUM OF THE TIME INCRE-MENTS BETWEEN SUCCESSIVE FAILURES, we multiply by an appropriate factor ϕ . Hence, the median waiting time becomes

$$_{(\frac{n}{r})}T_{.5} = \phi \, r^{1/b} \, \theta \left(\frac{\log_e 2}{n} \right)^{1/b} \tag{18}$$

An empirical expression for ϕ which has proven to be satisfactory in the author's experience is given by

$$\phi = \left[\frac{r - (1 - \ln 2)}{r \cdot \ln 2} \right]^{1/b} r! \; \frac{\Gamma \left(1 + 1/b \right)}{\Gamma \left(r + 1/b \right)} \tag{19}$$

It should be pointed out that the illustrations in this report are valid regardless of the true values of ϕ , since the charts 1 through 3 are independent of ϕ , and charts 4 through 7 are constructed by assuming the product $\phi L_{.5}$ = 400, without specifying any fixed numbers for ϕ or $L_{.5}$.

For n = r the median waiting time (18) is

$$_1 T_{.5} = \phi \, r^{1/b} \, \theta \left(\frac{\log_e 2}{r} \right)^{1/b} = \phi \theta \left(\log_e 2 \right)^{1/b} \tag{20}$$

Hence, the median waiting time ratio is

$$\rho = \frac{_{(\frac{n}{r})}T_{.5}}{_1 T_{.5}} = \frac{1}{\left(\frac{n}{r} \right)^{1/b}} \tag{21}$$

The relation (21) is plotted in charts 2 and 3 for Weibull slopes b = 1, 1.5, and 2.

The Answer to Question Number 3

Let Z = median total cost of the fatigue testing program. We break up the total cost into two components. These are:

91

1. Manufacturing cost of the specimens
2. Waiting cost (including operating expenses, cost of delay in putting a better product on the market, etc.).

The number of specimens which must be manufactured is $n + r - 1$, since immediately before the final failure there are n specimens running and $(r-1)$ have previously been failed, making a total of $n + (r-1)$ which must be available for the test. Therefore, TOTAL MANUFACTURING COST $= (n + r - 1)C_s$. (22)

$(C_s =$ manufacturing cost per specimen)

The median waiting time required to fail r specimens is given by (14):

Let C_w denote the waiting cost per hour, which we assume is known.

Therefore, MEDIAN TOTAL WAITING COST $= C_w \left[\left(\frac{n}{r} \right) T_{.5} \right]$

$$= \phi \, r^{\frac{1}{b}} \theta \left(\frac{\log_e 2}{n} \right)^{\frac{1}{b}} C_w = \phi \, r^{\frac{1}{b}} L_{.5} C_w \, n^{-\frac{1}{b}} \qquad (23)$$

$$\therefore \qquad Z = (n+r-1)C_s + \phi \, r^{\frac{1}{b}} L_{.5} C_w \, n^{-\frac{1}{b}} \qquad (24)$$

For $\left(\frac{n}{r} \right) = 1$, this total cost becomes $Z_1 = (2r-1)C_s + \phi \, r^{\frac{1}{b}} L_{.5} C_w \, r^{-\frac{1}{b}}$

Hence, the median cost ratio is:

$$\Omega = \frac{Z}{Z_1} = \frac{(n+r-1)C_s + \phi \, r^{\frac{1}{b}} L_{.5} C_w \, n^{\frac{1}{b}}}{(2r-1)C_s + \phi \, r^{\frac{1}{b}} L_{.5} C_w \, r^{-\frac{1}{b}}} = \frac{\left(\frac{n}{r} \right) + \left(\frac{r-1}{r} \right) + \phi \, r^{-1} L_{.5} \left(\frac{C_w}{C_s} \right) \left(\frac{n}{r} \right)^{-\frac{1}{b}}}{2 - \frac{1}{r} + \phi \, r^{-1} L_{.5} \left(\frac{C_w}{C_s} \right)} \qquad (25)$$

Let $\left(\frac{n}{r} \right) = \xi$ Then

$$\Omega = \frac{\xi + \left(\frac{r-1}{r} \right) + \phi \, r^{-1} L_{.5} \left(\frac{C_w}{C_s} \right) \xi^{-\frac{1}{b}}}{2 - \frac{1}{r} + \phi \, r^{-1} L_{.5} \left(\frac{C_w}{C_s} \right)} \qquad (26)$$

Let $\dfrac{1}{2 - \frac{1}{r} + \phi \, r^{-1} L_{.5} \left(\frac{C_w}{C_s} \right)} = \overline{K}$

Then $\Omega = \overline{K} \left[\xi + \left(\frac{r-1}{r} \right) + \phi \, r^{-1} L_{.5} \left(\frac{C_w}{C_s} \right) \xi^{-\frac{1}{b}} \right]$ (27)

92

The relation (27) is plotted in chart 4 for b = 1, and in chart 6 for b = 1.5. It is assumed that r = 20 and $L_{.5} = \dfrac{400}{\phi}$ hours.

The Answer to Question Number 4

To find the minimum value of Ω in (27), we differentiate with respect to $\mathcal{F}$ and then put $\dfrac{d\Omega}{d\mathcal{F}} = 0$.

Thus,

$$\frac{d\Omega}{d\mathcal{F}} = \overline{K} \left[1 - \frac{\phi}{b} r^{-1} L_{.5} \left(\frac{c_w}{c_s} \right) \mathcal{F}^{-1/b - 1} \right] = 0. \tag{28}$$

Solving (28) for the optimum value of $\mathcal{F}$ we obtain

$$\mathcal{F}_{opt.} = \left[\frac{\phi}{b} \frac{L_{.5}}{r} \left(\frac{c_w}{c_s} \right) \right]^{\frac{b}{b+1}} \tag{29}$$

Substituting (29) into (27) we obtain for the minimum cost ratio

$$\Omega_{Min.} = \frac{1 - \frac{1}{r} + (1+b)\mathcal{F}_{opt.}}{2 - \frac{1}{r} + b \, \mathcal{F}_{opt.}^{\frac{b+1}{b}}} \tag{30}$$

Relations (29) and (30) are plotted in chart 5 for b = 1, and in chart 7 for b = 1.5. It is assumed that r = 20 and $L_{.5} = \dfrac{400}{\phi}$ hours in both charts.

Conclusions

From charts 5 and 7 we conclude that whenever the waiting cost per hour is several times the manufacturing cost of a single test specimen, it is most economical to simultaneously run several times as many specimens as we are going to fail. In case a fatigue testing laboratory does not have sufficient capacity to accommodate large values of $\left(\frac{n}{F} \right)$, it would even pay to expand the test facilities to a point where large $\left(\frac{n}{F} \right)$ values can be handled. In the long run, the drastic reduction in total testing cost would soon pay for the cost of additional expansion. Furthermore, with large values of $\left(\frac{n}{F} \right)$ fatigue tests are completed in a fraction of the time it takes to complete them with small values of $\left(\frac{n}{F} \right)$. As a rule of thumb we can say that a large $\left(\frac{c_w}{c_s} \right)$ ratio requires a large $\left(\frac{n}{F} \right)$ ratio. On the other hand, if $\left(\frac{c_w}{c_s} \right)$ is very small, it can be seen that there is no justification for a large $\left(\frac{n}{F} \right)$. In short, the more expensive it is to wait, the more test facilities we need in order to operate at peak efficiency. This is nothing more than sound business judgment.

BIBLIOGRAPHY

1. Weibull, W., "A Statistical Distribution Function of Wide Applicability", <u>Journal</u> <u>of</u> <u>Applied</u> <u>Mechanics</u>, September, 1951, pp. 293-297.

2. Johnson, L. G., "The Median Ranks of Sample Values in Their Population With an Application to Certain Fatigue Studies", <u>Industrial</u> <u>Mathematics</u>, Volume 2, 1951, pp. 1-9.

3. Arkin, H. and Colton, R. R., "An Outline of Statistical Methods", <u>Barnes</u> & <u>Noble</u>, <u>Inc.</u>, 4th Edition, 1939, page 193.

4. Epstein, B., "Statistical Problems in Life Testing" (II p. 386) <u>Proceedings</u> <u>of</u> <u>the</u> <u>Seventh</u> <u>Annual</u> <u>Convention</u> <u>of</u> <u>the</u> <u>American</u> <u>Society</u> <u>for</u> <u>Quality</u> <u>Control</u>, May, 1953.

CHART 1
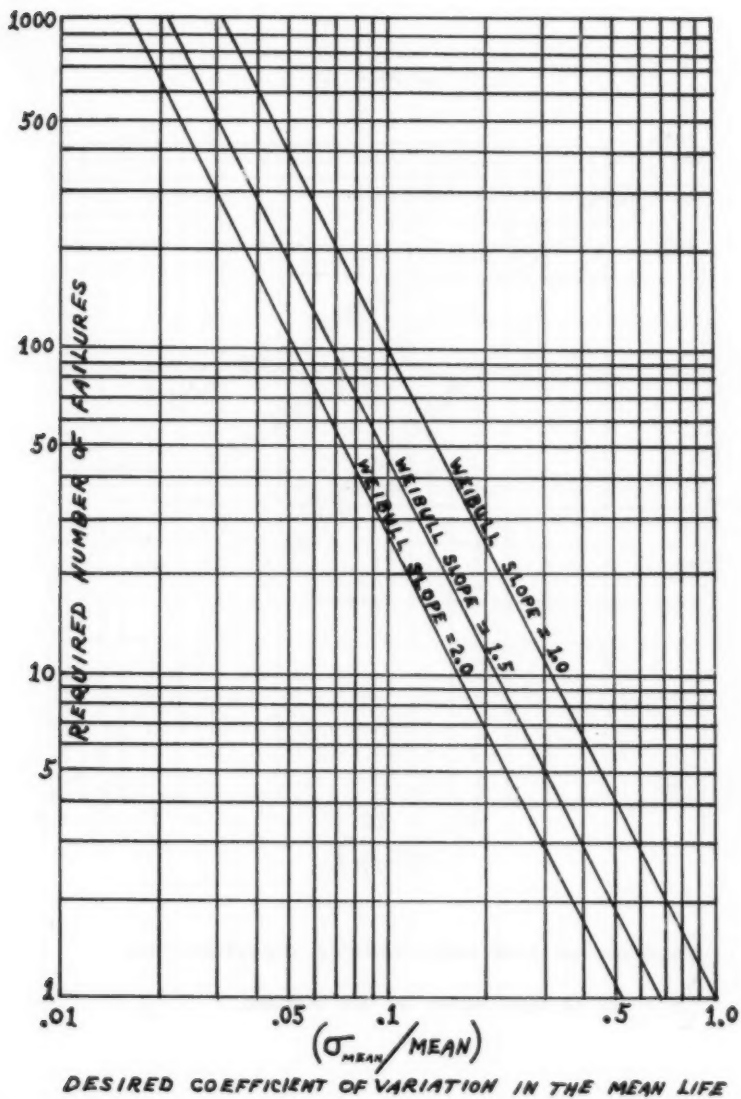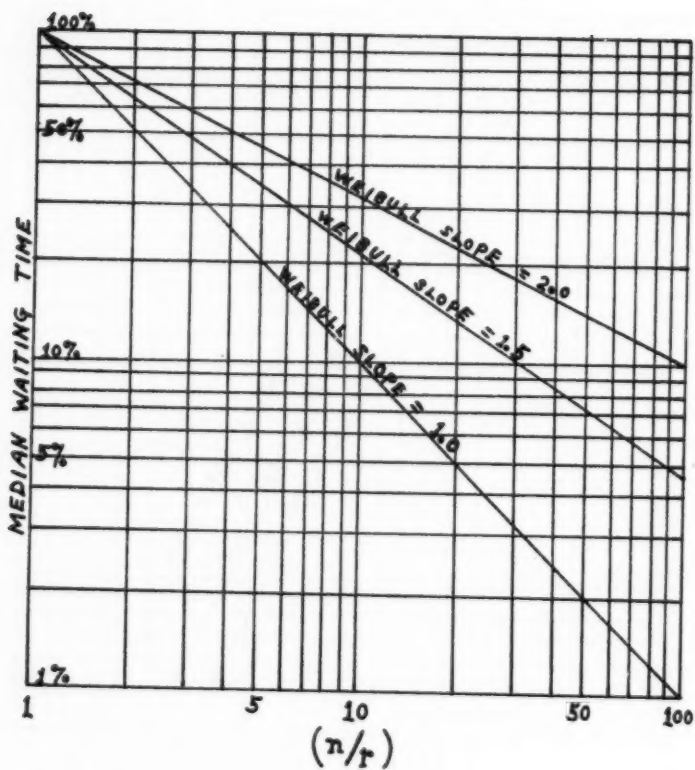
DESIRED COEFFICIENT OF VARIATION IN THE MEAN LIFE

CHART 2

n = NO. OF SPECIMENS RUNNING SIMULTANEOUSLY

r = NO. OF SPECIMENS TO BE FAILED

CHART 3

$$\begin{cases} n = \text{NO. OF SPECIMENS RUNNING SIMULTANEOUSLY} \\ r = \text{NO. OF SPECIMENS TO BE FAILED} \end{cases}$$

97
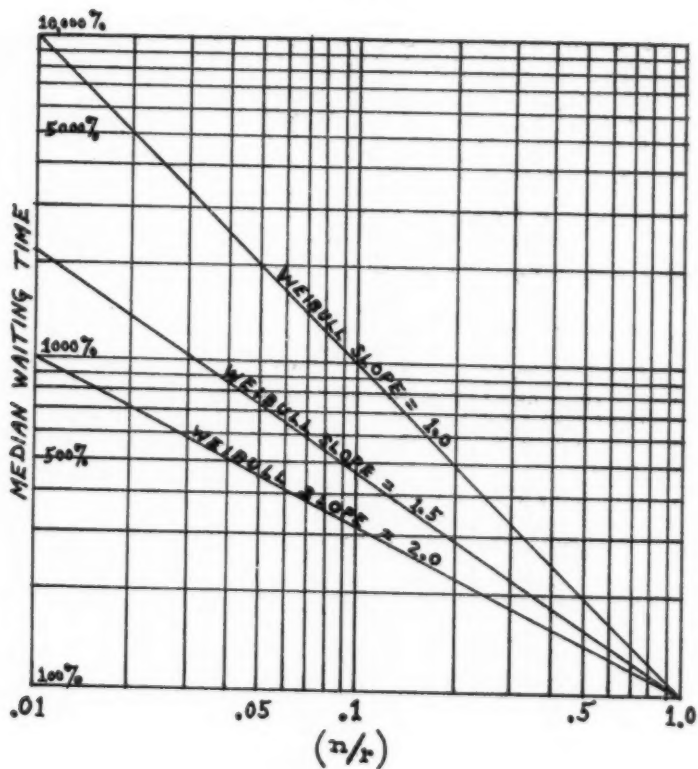
CHART 4

r = NO. OF SPECIMENS FAILED = 20

$$\text{MEDIAN LIFE} = \left(\frac{400}{r}\right) \text{HOURS}$$

WEIBULL SLOPE = 1.0

n = NO. OF SPECIMENS RUNNING SIMULTANEOUSLY

C_w = WAITING COST PER HOUR

C_s = MFG. COST PER SPECIMEN

CHART 5

r = NO. OF SPECIMENS FAILED = 20

MEDIAN LIFE = $\left(400/\phi\right)$ HOURS

WEIBULL SLOPE = 1.0
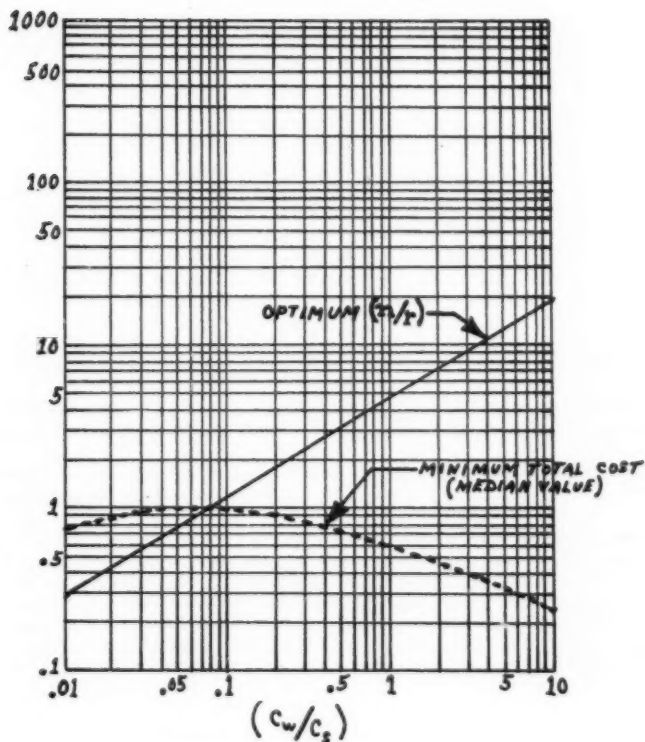
n = NO. OF SPECIMENS RUNNING SIMULTANEOUSLY

C_w = WAITING COST PER HOUR

C_s = MFG. COST PER SPECIMEN

CHART 6

n = NO. OF SPECIMENS RUNNING SIMULTANEOUSLY

C_W = WAITING COST PER HOUR

C_S = MFG. COST PER SPECIMEN

100

CHART 7

OPTIMUM (n/r)

MINIMUM TOTAL COST
(MEDIAN VALUE)

$\left(C_w/C_s\right)$

n = NO. OF SPECIMENS RUNNING SIMULTANEOUSLY

C_w = WAITING COST PER HOUR

C_s = MFG. COST PER SPECIMEN

101

THE USE OF QUALITY CONTROL TECHNIQUES IN OPERATIONS RESEARCH

R. W. Lindsay, Operations Research Supervisor
American Enka Corporation

This report covers an Operations Research project in which it was found desirable to lean quite heavily on techniques commonly associated with Quality Control. The project was a study of production and inventory control.

Work on this project was undertaken in the following sequence:

1. Formulate the problem and define its limits.

2. Develop a tentative model.

3. Test the model to see whether it is reasonable and make modifications where necessary.

4. Develop the information sources and practical measures necessary for actually using the model.

5. Determine whether use of the model accomplishes the purpose for which it was developed. This was done first by simulation and then by actual trial use.

6. Develop a procedure for ensuring that the model remains consistent with actual conditions in the future.

I believe you will see a lot of similarity between the above steps and those necessary to carry out a successful quality control project.

Before going into detail, I would like to give you a brief picture of the background of the project.

The rayon industry lived in an almost continuous sellers' market from its infancy in the Twenties to the post-war years of the late Forties. During this period, production and inventory control was relatively easy. As the industry came of age and moved to a buyers' market this control area became a headache. Our major problem took the form of frequent production program changes. The processes in a rayon plant are such that excessive program changes are very expensive in terms of cost, quality, and morale.

We were very interested in Operations Research, and since production and inventory control seemed to be a fertile field for this discipline, we decided to make the problem our first major project in Operations Research.

First we had to formulate the problem, define its limits and develop a trial model. In the initial phase, we seem to have departed from the path generally followed in this type of study. In most published studies on production and inventory control, the problem is defined primarily as that of obtaining an optimum balance between customer shortages, inventory and program changes. In our case, we were fairly well satisfied with our past performance in so far as inventory and customer shortages were concerned. It was also realized that the difficulty of

developing a realistic model of the comparative cost associated with these variables would greatly complicate the study. If we could reduce program changes without increasing either inventory or customer shortages, we would be certain that we were getting a real profit increase even without such cost figures. We, therefore, decided to make this our goal.

Examination of historical data on shipments showed large variations for individual products between time periods. The patterns were such as to lead us to suspect that a large part of the variation was of a random nature. However, it was obvious that significant changes in the true average market demand were superimposed on the random variations from time to time. We decided to try to design a probability model as an analysis tool to be used in differentiating between these two types of variation.

Since shipments are discrete occurrences in time, theortical considerations pointed to a Poisson distribution as the most likely one. However, some change of variable was necessary because the Poisson distribution deals with the number of discrete occurrences; whereas we were interested in the poundage involved in shipments, not in the number of shipments, as such.

The problem is similar to that of analyzing the number of cars involved in highway accidents. The number of accidents occurring may fit the simple Poisson, but the variation in the number of cars involved per accident makes a compound distribution necessary. Our problem was more difficult because the possible variations in poundage per shipment are almost infinite.

We tried fitting a Poisson curve to our shipment data, using an average weight per shipment. This attempt was a complete failure. Investigation showed that the small percentage of large shipments, by their occurrence or non-occurrence in a period of time, caused the actual data to vary much more than our model. However, the general shape of our histograms still suggested Poisson.

We then designed a shipment function for each product line which would give increased effect to the larger shipments. The function was calculated by dividing the summation of the squares of the individual shipment weights by the summation of the individual weights.

Such a function can be justified theoretically. It is generally accepted that the sum of several variables, with the same basic distribution, will result in a distribution with a variance equal to the sum of the variances of the individual distributions. The variance of the distributions based on our shipment function meet this criteria.

A further check on the validity of the basic model was made by comparing the resulting shipment distributions, for a representative group of products, with the corresponding Poisson distributions by means of a chi-square test. A good fit was obtained in most instances. In the instances where the fit was poor, it was obvious that the actual data was not all from a common distribution. In other words, a change in demand level had occurred for the product concerned during the time period covered. Figure 1 shows a typical histogram comparing the actual and theoretical distributions.

We also found that the calculated shipment function varied with the average demand level. Luckily, this was found to be a straight line function, and we were able to establish workable relationships for various product groups from existing data. There was some interesting development work involved. When samples are drawn from a two-dimensional distribution, a scatter diagram of elliptical shape is formed, assuming normal distributions. There are three regression lines which may be calculated from the data. (See Figure 2.) We have the regression of y on x, when y is the dependent variable and x is the known variable. This line is useful in estimating y from a knowledge of x. We also have the regression of x on y, when y is the known variable. In addition, we have a third line which represents the structural relationship between the variables. Such a line is sometimes called a weighted regression line. One situation where this line is needed occurs when it is not practical to estimate either variable separately but the two must be estimated jointly. When shipments for a product go out of limits, changing the demand level means simultaneously changing the shipment function. Therefore, with both variables unknown, we are interested in the structural relationship.

Using these variables in combination with a Poisson distribution, we were able to construct realistic models of shipment activities. From these models, minimum and maximum shipments were determined throughout the volume ranges of our product lines. Figure 3 shows a chart for a typical product line. The center line represents the true average demand. The other two lines represent the limits of the expected random fluctuations about this average, for a predetermined level of confidence. It was necessary to construct such charts originally in terms of the shipment functions and then convert the scales to pounds. It will be noticed that the variation, as a percentage of the average, is much larger for small values of the average than for large ones. This is typical of a Poisson distribution. The limits are used as control limits to determine when a significant change in the demand level takes place. The limits also serve as a measure of the inventory required to cover normal shipment variations from inventory without a program change.

I mentioned earlier that our objective was to reduce program changes without increasing customer shortages or inventory. We found that limits based on a five per cent level of confidence came very close to perpetuating the overall levels of inventory carried in the past.

At this point in our work, the development and testing of the basic model was complete. The next job was to develop the practical measures for actually using the model.

A major problem was encountered in determining the time period over which to analyze shipments. On a weekly or bi-weekly basis, only very large changes could be detected because of the large random variations. If a longer time period was used, the changed level could seriously affect inventory before it had been in effect through even one complete time period.

We finally decided to organize data weekly and to compare the data for the latest periods of one, two, four, eight, and sixteen weeks with the proper limits. In this way the larger changes were discovered quickly by means of limits for the shorter periods. Small changes or those occurring gradually over a period of time were discovered by means of the

limits for the longer time periods.

We also had another unique problem. It was not sufficient merely to determine that the demand level had changed for a product, it was also necessary to obtain a useful estimate of the new demand level and this, to be effective, had to be done on the basis of one sample from the new population.

We adopted the expedient of shifting the demand level just enough to make the revised limits include the out-of-limit value, as illustrated in Figure 3. In many cases this underestimates the magnitude of the change initially but eventually values close to the true limits of the new population will indicate accurately the true demand level. In the meantime, we would be in little danger of trouble with no extreme values occurring.

I have talked mainly about shipments so far. However, we found it very beneficial to apply this technique to orders received as well as shipments. Shipments and orders behave very much as independent variables with the same average. Orders are received some time before shipment is made, as a rule. The weight function is also different for orders, due to the tendency for customers to specify more than one shipment on an order. We found that a separate analysis of shipments and orders almost doubled our chances of detecting real changes in demand level over the chances using shipments alone.

Using control limits in the manner described made it possible for us to determine the limits within which shipments would probably fall in a future planning period assuming no specific knowledge about that period. However, these limits were rather broad, and we usually have a valuable piece of information about future shipments; namely, unfilled orders on hand. It seemed logical to make use of this information to narrow the limits.

Since unfilled orders represent a partial knowledge of future shipments, taking them into consideration represents a problem in conditional probabilities.

Let me draw a parallel:

You wish to know the minimum and maximum number of heads you are likely to obtain by tossing 20 coins. Based on a five per cent level of confidence, you would expect between 6 and 14. However, let's assume you are tossing the coins one at a time. You have already tossed 10 of them and know this partial result. The minimum and maximum number to be expected from this particular trial is therefore the number obtained from the 10 already tossed, plus the minimum and maximum number to be expected from the remaining 10 coins, which is between 2 and 7. Thus the spread from minimum to maximum for this particular trial is only 6 when the result of half the trial is known as compared to 9 without this information.

We applied this principle in the case of the unfilled orders. The average order lead time was determined for each type of product. This lead time was then subtracted from the planning period and the maximum shipments determined for the remaining period. We call this figure the conditional component. The unfilled orders plus the conditional com-

106

ponent then gives us the maximum shipments for the planning period.

We were now ready to test the practical usefulness of the model. The completed Production and Inventory Control Plan was applied "on paper" to a year's data. Paper programs were determined and resulting paper inventories were calculated using actual order and shipment data. The results as to average inventories, machine changes, and customer shortages were then compared with actual results for the same period of time using methods of control then in effect. The control plan reduced both machine changes and customer shortages by about fifty per cent while overall average inventories during the year were different by less than two per cent.

Production programs had been determined in the past by the Sales Division on the basis of outstanding orders and current inventories conditioned by information obtained from customers as to their future buying intentions.

The Production and Inventory Control Plan was instituted for a trial period of six months, with the Sales Division retaining a veto power on the basis of their market information. The six months' experience was satisfactory. The Control Plan was then adopted and was turned over to the Production Planning Department, to be used on the same basis as during the trial period. Experience thus far is fairly well in line with predictions made as a result of the original application "on paper."

Plans were made to audit the control system periodically in order to keep the parameters of the model up to date. We are now engaged in the first of these audits. In the course of the audit, we are also re-examining the whole structure for possible improvements. For instance, our original model is based on an overall inventory in line with past experience. We are now attempting to build into the model an optimum inventory concept which would minimize the total program change costs and inventory carrying costs.

I have covered only the broad outlines of our development work. Many problems of a lesser nature were encountered both in developing the basic shipment and order models and in designing means for using these models in preparing a detailed production program. Solutions to many of these problems were derived through simulation.
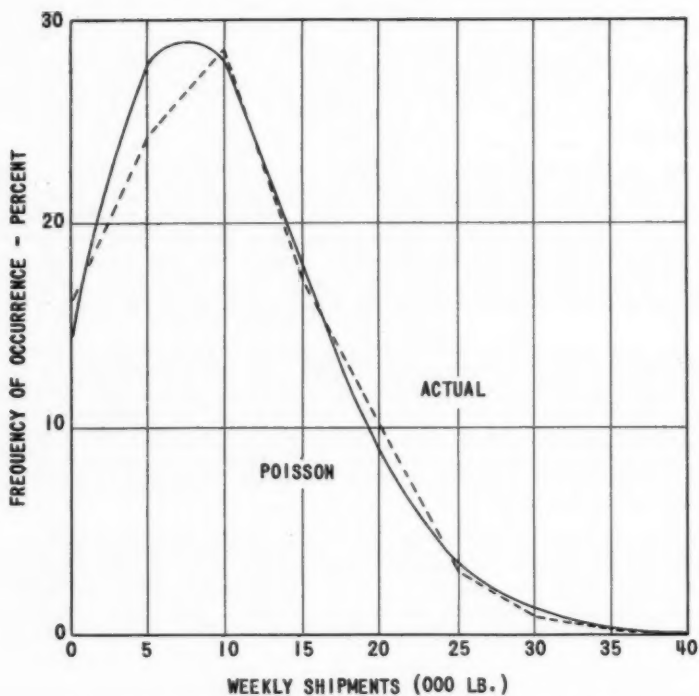
FIGURE I - ACTUAL AND THEORECTICAL DISTRIBUTION OF SHIPMENTS FOR A MEAN OF 10,000 LBS. PER WEEK, AND A SHIPMENT FUNCTION OF 5,000 LB.
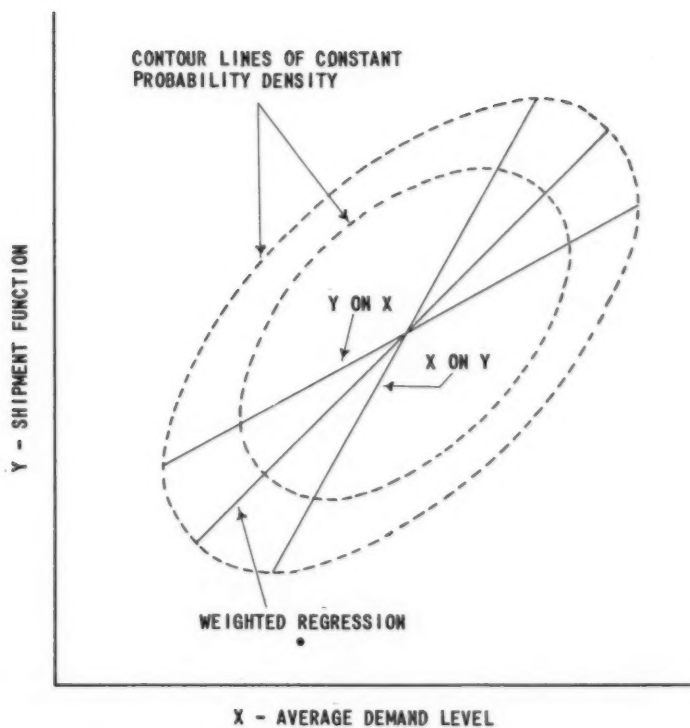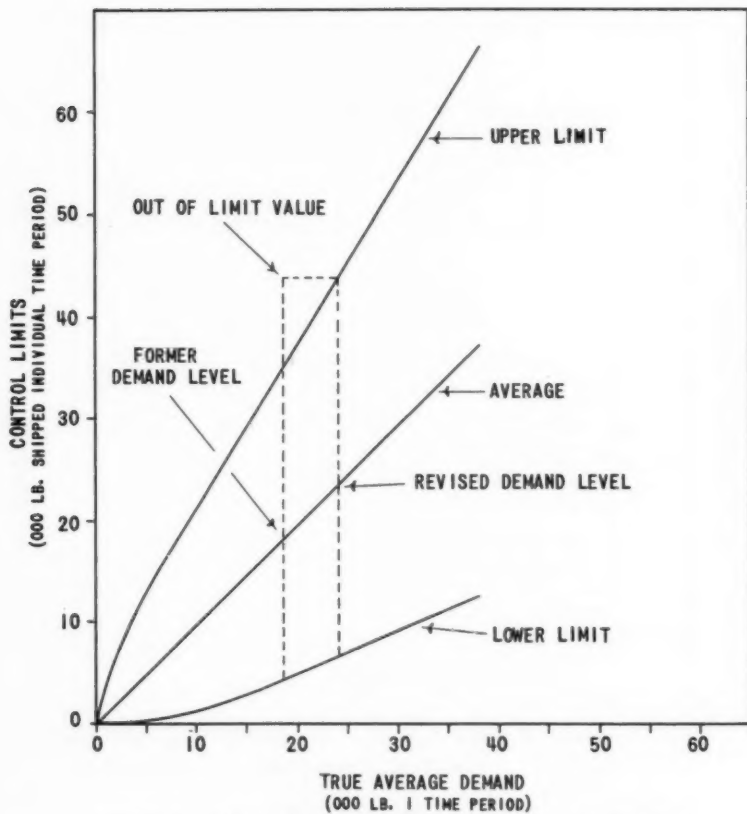
108

FIGURE 2 - ALTERNATIVE REGRESSION LINES

109

FIGURE 3 - TYPICAL RELATION BETWEEN AVERAGE DEMAND AND
CONTROL LIMITS

110

COMPENSATING FOR SYSTEMATIC EFFECTS DETECTED
BY AN ANALYSIS OF COVARIANCE IN A MICROBIOLOGICAL ASSAY

Norman F. Knowlden
Lederle Laboratories Division, American Cyanamid Company

The Situation:

The basic methods for measuring the antibiotic content of a given
material are microbiological. In this type of method, the inhibition
which the material causes on the growth of a particular microorganism is
compared to the inhibition displayed by different levels of potency of
a standard antibiotic. One of these microbiological assay methods is
known as a "turbidimetric" method. In this assay method, the amount of
growth is determined by measuring the amount of turbidity which results
from the concentration of microorganisms present. A filter photometer
is used to measure the light blocking effect of this turbidity. A
photocell picks up the light passing thru the sample and a galvanometer
shows the intensity of this light. Thus the higher the galvanometer
reading the greater the antibiotic potency and a typical growth-response
curve results, (Slide 1).

The general method (Slide 2) is to dilute both unknown and known
samples to a level of potency where the change in growth rate is most
sensitive to a change in antibiotic potency. This level is specific for
every combination of antibiotic and microorganism resulting in inhibition.
The diluted samples are placed in duplicate tubes. Into each tube is
then placed an inoculum containing a medium in which the organism can
flourish and a heavy concentration of rapidly growing organisms all being
at about 37°C. The assay tubes are next placed in a 37°C. water bath
incubator for 3 hours. Following this, the whole "test" is "killed" by
the addition of a dilute formaldehyde solution to every tube.

In the method studied, the standard antibiotic and the unknown sam-
ples, based upon the predicted potency, were diluted to a potency that
would give the desired amount of antibiotic in the assay tube. Then
five subsequent dilutions were made of this and placed in assay tubes,
in duplicate. Thus each sample was represented by 10 tubes in the assay.
An automatic pipetting machine was then used to add the inoculum. This
was added to the tubes in each row in order and to each rack in order so
as to maintain an order throughout the whole test. The assay was then
incubated 3 hours after which it was "killed" by the addition of formal-
dehyde solution to every tube following the same order as that used for
inoculation. Turbidity was then determined in a Lumetron Colorimeter.
The galvanometer readings of duplicate assay tubes were averaged and the
results "read" from standard curves prepared from standards run in the
same way at the same time.

The Problem and Its Analysis

Now the problem as presented to me was routine. The laboratory had
been running check samples, i.e. samples of the standard compared against
itself, for the past two months. These potency results were presented
to me and I was asked to calculate confidence limits for the assay. This
is a standard type of calculation described in many texts.

Upon examining the data, it was obvious without testing that there

111

SLIDE 1

TYPICAL GROWTH-RESPONSE CURVE

GALVANOMETER READINGS

INCREASING ANTIBIOTIC POTENCY →
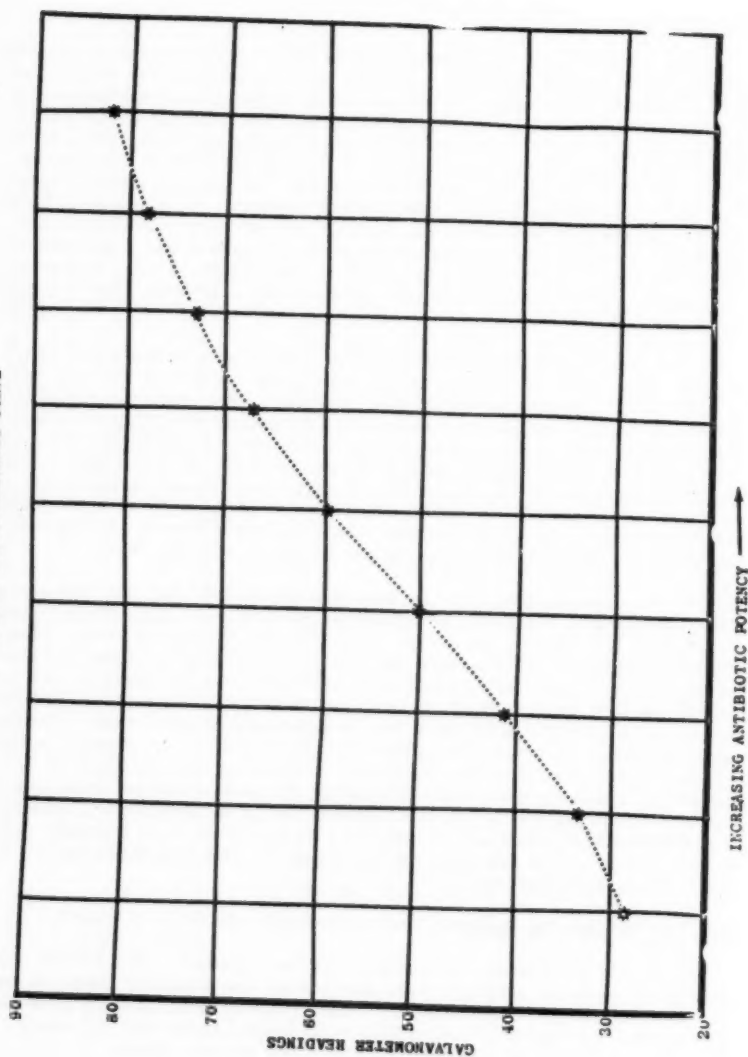
SLIDE 2

FLOWSHEET OF ASSAY PROCEDURE
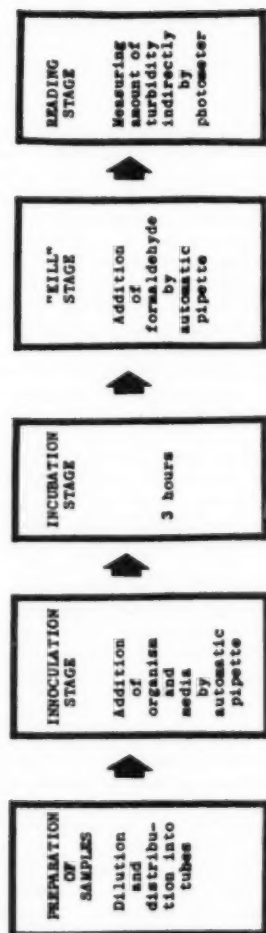
was a drastic difference between the two months. Being samples of the standard one would expect the average potency found to be 100%. This was not the case. The first month's samples averaged 92% and ranged from 80% to 95% whereas the second month assay average was 99% and ranged from 95% to 103%. Why should these two months differ so? What was the cause of this bias? This was the real problem, not the calculation of confidence limits.

This situation was discussed with the laboratory supervisor and two routine suggestions were considered as possible answers for the difference between months. First we have the age-old possibility that the technicians had become aware they were assaying samples of the standard and were using "extra care" and secondly we have the possibility that sufficient experience with the assay had brought about this improvement.

Neither of these possibilities seemed to fit the circumstances since both would affect precision rather than bias. Consequently I asked to see a routine assay being run. As stated before the standard and unknown samples were prepared to the appropriate levels and placed in duplicate tubes. These tubes were then placed in a rack in the manner shown. (Slide 3). The duplicates were always adjacent to each other with the tubes of any one sample common to one column in a rack. Each rack contained 4 complete samples, and 20 racks was the usual size of an assay. The routine procedure was to assign to the first rack tubes containing blanks which would be used to adjust the instrument. Following this would be the racks containing standard solutions for preparing a daily standard curve and following these would be the racks containing the unknown samples, each sample being completely contained in one single column of one single rack.

The tubes were then inoculated (see arrows) proceeding from rack 1 to rack 20. The racks were then placed in the incubator using a standard arrangement (see diagram). After incubation the "kill" stage was carried out in the same order as inoculation and then the sample turbidity read in order as the racks became ready after "kill."

This systematic procedure proved to be very efficient as far as handling the tubes was concerned but like all systematic procedures it lent itself to introducing bias in the assay. As a matter of practice always examine how your data was collected. Many times explanations of a problem will be apparent and only in this way can appropriate experimental designs be developed.

To substantiate my opinion that this systematic procedure might be the cause of our trouble, I requested that some experimental work be done. This was agreed to provided it could be arranged so that it did not interfere with the normal rate of output of assay results. Even though it was acknowledged that a bias may be present, it was recognized that this bias appeared to be such that it provided extra assurance that the true potency results were at least as potent as reported if not more so. The urgency of results, even if biased, was necessary to allow other development studies to continue. In production work this is quite often the case. Rarely can the most appropriate experimental design be employed without interfering with the production schedule.

114

SLIDE 3

BEFORE

ARRANGEMENT OF RACKS AND TUBES WITH ORDER OF

INOCULATION, "KILL", ETC. PRIOR TO STUDY
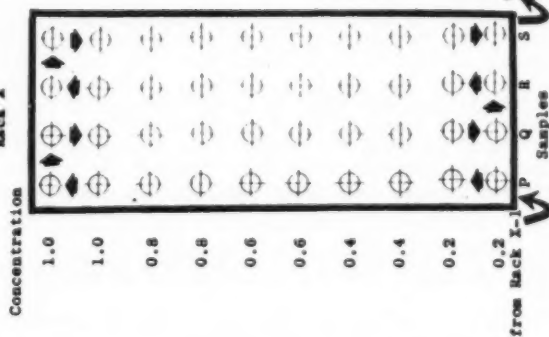
POSITION OF RACKS IN INCUBATOR

1	5	9	13	17
2	6	10	14	18
3	7	11	15	19
4	8	12	16	20

115

My first request was that in the next assay, duplicate tubes of each of two concentrations of the standard, be included in each rack used, and the routine procedure be followed. Where these were placed in each rack was of little consequence as I was first interested in showing that rack differences existed. This experiment was agreed to as it was felt that not many samples would be deleted. The data was collected and analyzed as in Table I. Graphically the galvanometer readings were plotted as in slide 4. The circles represent individual readings while the astericks are the average of each pair of duplicates. It is evident here, as in the analysis of variance, that there is a highly significant difference between racks. Since the tubes of each sample was contained within one rack whereas the tubes of the standard curve was always in another rack, these rack differences evidently were the cause of our bias. Low galvanometer readings for samples would result in low potencies.

Looking at the plot of results there appears to be a nice downward linear trend with the order of processing. Consequently the processing order was considered as another variable and an analysis of covariance performed on the same data as shown in Table I. What this does is to remove the effect of the processing order from the measurement of rack differences or graphically we now consider rack differences as whether there is any significant difference in the amount of deviation from this processing order regression between the racks. The racks were now found not to be significantly different. We now had our answer, the processing order could introduce bias. But how could this explain the bias found in the original data. Investigating the records kept over the time the original data was collected it was found that during the first month the check samples were always assigned to the last rack used whereas during the second month they were always assigned to the rack immediately following the standard. This was our answer. Also it was realized that in all past assays there existed an increasing negative bias from the first to the last rack processed each day.

This is all right to know but how do we go about eliminating this processing order effect so that future potency calculations will not be affected? A correction term for each rack could be developed each day but the calculations would be impractical for routine work. Randomization was out of the question. Imagine randomizing 800 tubes. The extra time involved and means of identification of tubes would be tremendous. The possibility of confusion which would lead to complete loss of a days work was too great. Also unless several replicate tubes were used, randomization would only transfer the rack to rack variation to variation between duplicates and our resulting precision would be poor.

Considering the routine that was being followed, it was reasoned that this systematic ordering could produce an effect in several places.

1. During inoculation there was a time difference of 10 minutes between when the first rack was inoculated and when the last rack was inoculated.

2. The position in the incubator could be important if the circulation were not uniform.

3. The order of kill could be important since there again was a

116

10 minute lag between the first and last rack killed.

4. The order of reading could be important if there were a drift in the photometer with time.

5. Some combination of these might be causing the trouble.

Discussing these possibilities with the personnel involved, it was decided that the incubator position was the most likely possibility. The 10 minute lag during inoculation and kill was not thought to be important when you consider the tubes are in the incubator for 3 hours.

Since the results of the first experiment were so definite, further experimenting was agreed to provided it would not interfere with production. We considered it best to investigate one factor at a time rather than design an overall experiment. It would take more time of course and would be troublesome if interacting factors were the cause of our trouble but we felt there would be less chance of confusion. Also the technicians could understand what we were trying to do and should we find a solution to our problem there would be little difficulty in installing any changes necessary. This is of prime importance in production studies. Unless an improvement, when found, can be implemented the study will be just a waste of time.

Considering incubator position, a second experiment was run maintaining the usual routine except for placing the racks in the incubator. This was done randomly, using random numbers. Analyzing the results of this experiment, similar results as before were found (Slide 4). Processing order independent of incubator order was still significant. An analysis of covariance as in Table II verified this. Also an analysis of variance on the row and column positions in the incubator showed the incubator positions as not being important.

Next the order of inoculation was considered. This was done by randomizing the order of inoculation of racks while maintaining the systematic order in the remainder of the assay. It might be well to mention here that at no time will randomization within racks be considered for reasons stressed before. We felt that the systematic bias between racks was of greatest importance and would worry about within racks later. The data on inoculation, when collected and examined, again presented a similar picture. Analysis of covariance showed processing order independent of inoculation order to still be the culprit. (Table III).

Continuing we next randomized the order of kill from rack to rack maintaining the routine procedure in all other parts of the assay. Analysis of the results still showed rack differences but of a different nature than before. (Slide 5). An analysis of covariance on processing order no longer showed improvement. (Table II). There was no significant processing order effect when independent of the order of kill. If we performed an analysis of covariance on order of kill independent of the routine order of the rest of the assay we could explain the rack differences. Graphically this picture, where we consider order of kill instead of processing order, would have a picture similar to before (Slide 4). Here was the root of our trouble, the order of kill.

To round out all factors, the order of reading was next randomized

117

SLIDE 4

ROUTINE PROCEDURE
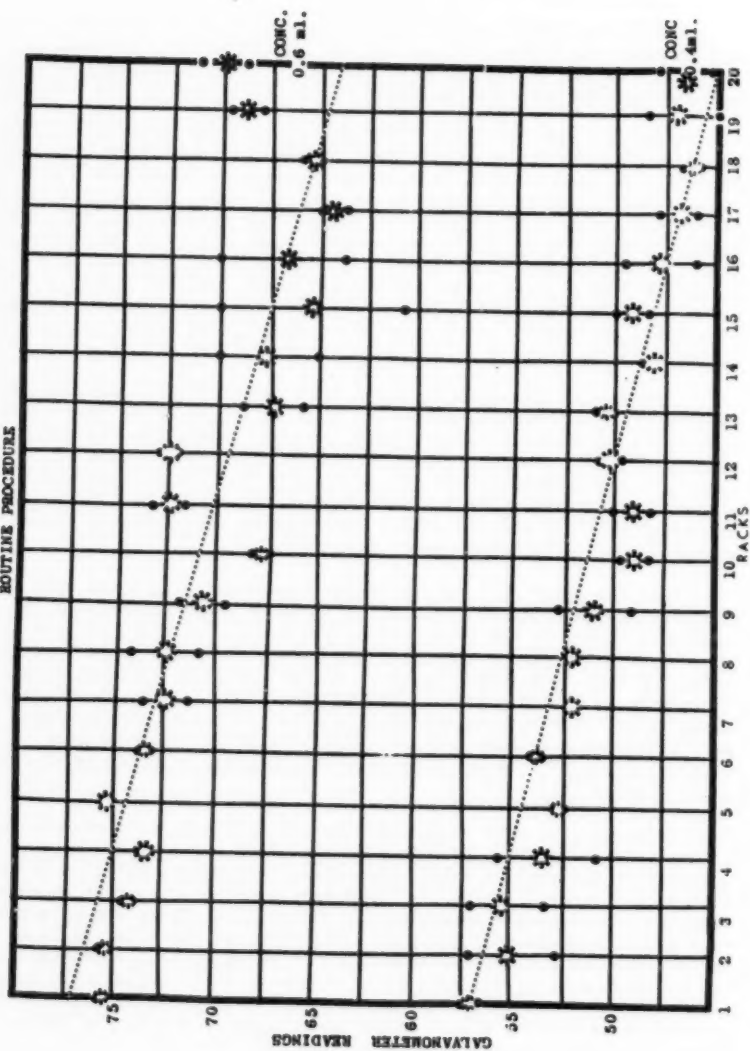
SLIDE 5

RANDOMIZED ORDER OF "KILL"
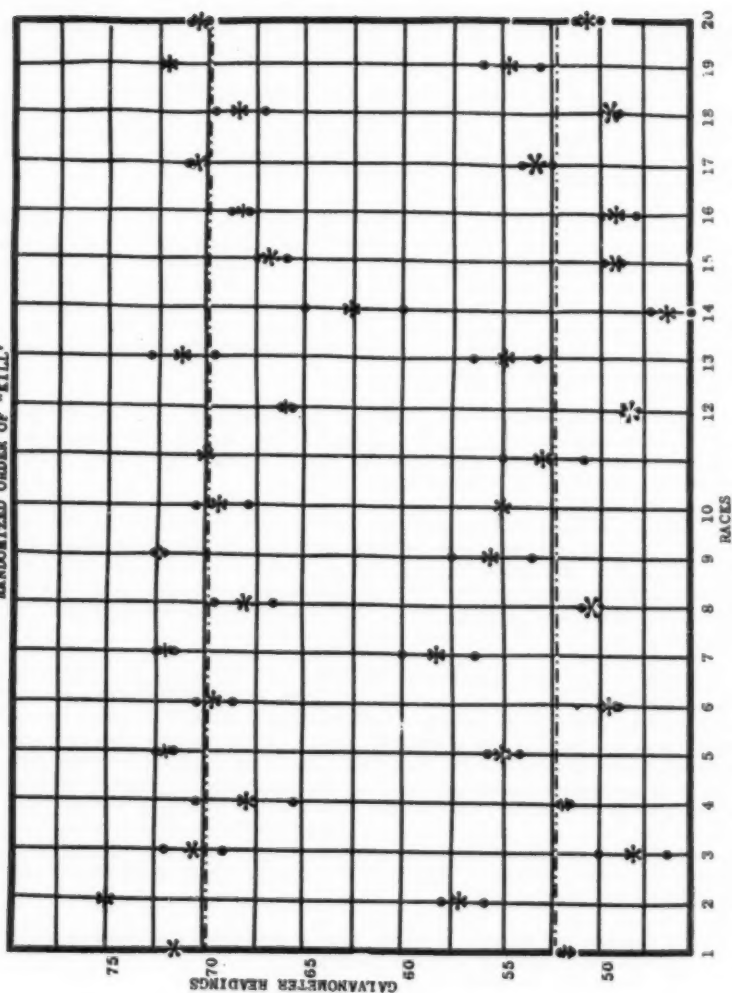
and found to be of no importance as a contributing factor.

As a result of these experiments, it was concluded that the order of kill was the major factor of interest. Why should this be so? It was realized that once started, the length of time an organism continued to multiply was not only during incubation but also during the period between removal from the incubator and the addition of the formaldehyde. This could be from 0 to 10 minutes progressing steadily from the first to the lask rack killed. This was at a time when growth was progressing at a rapid rate whereas during inoculation the growth was progressing at a slow rate.

How might this time difference be equalized from rack to rack? If our reasoning is correct, this is what is necessary to produce a suitable assay. This time equalization must be done in a manner that will lend itself to routine usage. There is no sense finding a solution if we can't implement it.

Let us return to the set-up used at the start of the assay (Slide 3). We have duplicate tubes adjacent to each other in each rack. What purpose do they serve? They simply provide a measure of variation of the technician and machine from tube to tube. This is of little importance to us as far as assay error is concerned. Samples are contained in different racks so rack differences are of principal importance when considering assay error. It is rack differences that must be minimized as much as possible.

The procedure after the tubes are prepared, is to inoculate these tubes using an automatic pipetting machine so the time interval is constant between tubes. The killing step also has equal time intervals since an automatic pipette is used.

Now if we consider instead of inoculating or killing in the routine manner, inoculating one tube of each pair of duplicates starting from the first rack and going through the last rack and then return inoculating or killing the second of each pair from the last rack through the first rack, the average time of inoculation and the average time of kill will be the same for each pair of tubes.

This can be made clearer if we order our tubes in the racks in the manner of slide 6. This procedure will not only equalize the time from rack to rack, the point we have been trying to achieve, but also will equalize the time for each pair of tubes. If our reasoning is correct this should minimize within rack differences as well as between rack differences thereby eliminating all bias since the average time during which growth can occur will be the same for all pairs of tubes. Of course the variation between the so called duplicate tubes will increase. This source of variation can then be corrected for the revised processing order effect by analysis of covariance to give an estimate of technician precision.

If successful we will wind up with a systematic procedure which
1) is suitable for routine purposes
2) requires no extra work
3) has simple potency calculations from the sample reading averages
4) has little likelihood of being confusing

SLIDE 6

AFTER

ARRANGEMENT OF RACKS AND TUBES WITH ORDER OF INOCULATION, "KILL", ETC. AFTER STUDY
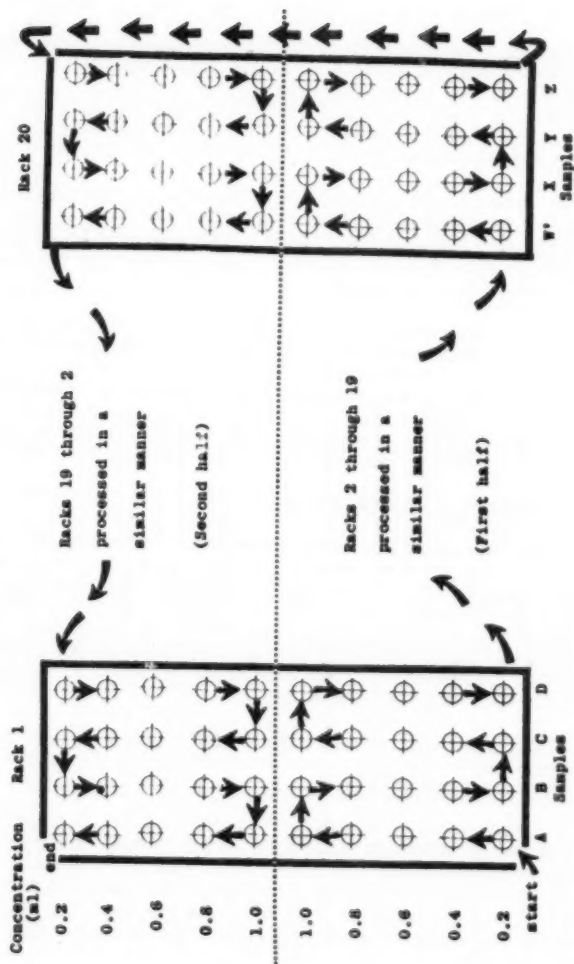
5) and provides a valid estimate of the assay error.
This procedure was tried first by using the revised ordering during the kill step only. The results on analysis are in Table VI. Graphically they plotted as in slide 7. There is no longer any downward trend of the averages and there is no significant differences between racks. This was exactly what was desired. All rack to rack bias was eliminated.

It was then decided to consider this revised ordering for both inoculation and kill. This would provide a systematic ordering from the beginning to the end of the assay. The results of this experiment were so promising that a second verification run was made and the experimentation concluded. Slide 7 is a graphical picture of the results of one of those experiments.

To estimate technician precision an analysis of covariance with the revised processing order was performed. Essentially this is done by determining the variation of each individual measurement from a linear regression line passing in order from the first to the last tube processed at each concentration.

In all cases the magnitude of the error between duplicate tubes, corrected for the revised processing order, was no greater than the concentration by rack interaction originally present. This corrected error term was now a valid estimate of the assay error and there was no longer any significant bias to affect potency calculations. It was now a simple matter to calculate the confidence limits originally requested. The 95% confidence interval was found to be $\pm 5.5\%$.

Implementation of the Solution:

With six operators to be taught the new technique, supervision expected to have to throw away a few days work, at least, before a successful assay could be completed. However, after the people had been given a full explanation of why the change was being made, the revised procedure was put into regular operation the first day. From then on it was a success. Those of you who are acquainted with microbiological assays will agree that a 95% confidence interval of $\pm 5.5\%$ is extremely good for this type of method.

This technique has since been applied to other turbidimetric microbiological methods in the same laboratory and the application has been completely successful.

At this time I would like to cite a remark the well known statistician, R. L. Anderson, made at a conference in November, 1956; "If the experimental units do differ widely on some measurable characteristic, x, it may be desirable to use a balanced assignment plan, so that the average value of x is the same for each treatment, and then use an analysis of covariance to obtain an unbiased estimate of the error variance. I predict that you will hear more of this procedure in the future." The study I have discussed was made in January, 1956 and we had been using this procedure for almost a year before Anderson made his remark. You will hear more of this technique in the future!

122

SLIDE 7

REVISED PROCEDURE OF INOCULATION AND "KILL"

123
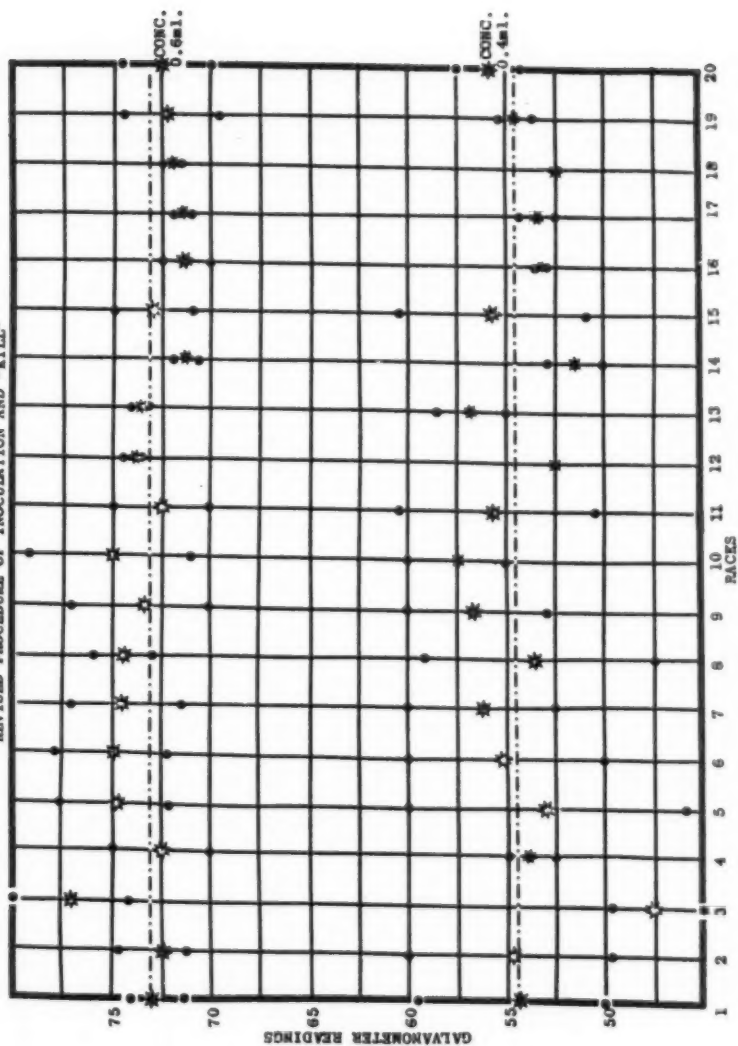
References:

1) Knudsen, L. F., "Statistics in Microbiological Assay", Annals of the N. Y. Acad. of Sciences, v. 52, Article 6, p. 889, (March 10, 1950).

2) Bennett, C. A. and N. L. Franklin (1954) <u>Statistical Analysis in Chemistry and the Chemical Industry</u>, John Wiley and Sons, New York.

3) Anderson, R. L. "Complete Factorials, Fractional Factorials and Confounding," p. 51, unpublished hand-out for Industrial Experimental Design Conference, November 5-9, 1956, at Institute of Statistics, North Carolina State College.

TABLES OF ANALYSIS OF VARIANCE

Table I

Ia.

Source	D.f.	S.S.	M.S.	F
Concentrations	1	7960.05	7960.05	10.2**
Racks	19	824.82	43.41	
Conc. X Racks	19	83.07	4.37	1.04
Duplicate Tubes	40	168.25	4.21	
Pooled Error	59	251.32	4.26	

Table II

IIa.

Source	D.f.	S.S.	M.S.	F
Concentrations	1	7440.16	7440.16	10.5**
Racks	19	1656.78	87.20	2.28*
Conc. x Racks	19	158.28	8.33	
Duplicate Tubes	40	145.88	3.65	
Columns	3	309.45	103.15	1.39
Rows	4	455.06	113.77	1.53
Rows x Col.	12	892.27	74.36	8.93**
Racks (Total)	19	1656.78		

TABLES OF ANALYSIS OF COVARIANCE

Ib.

	Sy^2	Sxy	Sx^2	D.f.	S.S. Adjusted	M.S.	F
Concentrations	7960.05	0	0	1	7960.05	7960.05	1.41
Racks	824.82	-345.25	166.25	18	107.84	5.99	
Conc. X Racks	83.07	0	0	19	83.07	4.37	1.04
Duplicate Tubes	168.25	0	0	40	168.25	4.21	
Pooled Error				59	251.32	4.26	

IIb.

	Sy^2	Sxy	Sx^2	D.f.	S.S. Adjusted	M.S.	F
Concentrations	7440.16	0	0	1	7440.16	7440.16	1; 1.97
Racks	1656.78	-503.94	166.25	18	129.23	7.18	2.28*
Conc. x Racks	158.28	0	0	19	158.28	8.33	
Duplicate Tubes	168.25	0	0	40	145.88	3.65	

** Significant at p = .01; * Significant at p = .05

TABLES OF ANALYSIS OF VARIANCE AND COVARIANCE

Table III

Source	D.f.	Sy²	M.S.	F	IIIb. Sxy	Sx²	D.f.	S.S. (Adjusted)	M.S. (Adjusted)	F	IIIc. Sxy	Sx²	D.f.	S.S. (Adjusted)	M.S. (Adjusted)	F
			IIIa.													
Conc.	1	6900.61	6900.61		0	0	1	6900.61	6900.61		0	0	1	6900.61	6900.61	
Racks	19	803.08	42.27	4.9***	-283.13	166.25	18	320.92	17.83	2.07	-61.38	166.25	18	780.42	43.36	5.0**
Co.xRa.	19	164.01	8.63	1.79	0	0	19	164.01	8.63	1.79	0	0	19	164.01	8.63	1.79
Du.Tub.	40	193.25	4.83		0	0	40	193.25	4.83		0	0	40	193.25	4.83	
Pld.Err.	59	179.64	3.04				59	179.64	3.04				59	179.64	3.04	

Table IV

Source	D.f.	Sy²	M.S.	F	IVb. Sxy	Sx²	D.f.	S.S. (Adjusted)	M.S. (Adjusted)	F	IVc. Sxy	Sx²	D.f.	S.S. (Adjusted)	M.S. (Adjusted)	F
			IVa.													
Conc.	1	6248.11	6248.11		0	0	1	6248.11	6248.11		0	0	1	6248.11	6248.11	
Racks	19	623.45	32.81	8.1**	-94.13	166.25	18	570.15	31.68	7.8**	-289.25	166.25	18	120.20	6.67	1.7,2.2*
Co.xRa.	19	76.89	4.05	1.58	0	0	19	76.89	4.05	1.58	0	0	19	76.89	4.05	1.58
Du.Tub.	40	102.75	2.57		0	0	40	102.75	2.57		0	0	40	102.75	2.57	
Pld.Err.	59	179.64	3.04				59	179.64	3.04				59	179.64	3.04	

Table V

Source	D.f.	Sy²	M.S.	F	Vb. Sxy	Sx²	D.f.	S.S. (Adjusted)	M.S. (Adjusted)	F	Vc. Sxy	Sx²	D.f.	S.S. (Adjusted)	M.S. (Adjusted)	F
			Va.													
Conc.	1	7144.20	7144.20		0	0	1	7144.20	7144.20		0	0	1	7144.20	7144.20	
Racks	19	1356.26	71.38	25.8**	-463.13	166.25	18	66.13	3.67	1.32	-19.13	166.25	18	1354.06	75.23	27.2**
Co.xRa.	19	48.43	2.55	1	0	0	19	48.43	2.55	1	0	0	19	48.43	2.55	1
Du.Tub.	40	155.00	2.88		0	0	40	155.00	2.88		0	0	40	155.00	2.88	
Pld.Err.	59	163.43	2.77				59	163.43	2.77				59	163.43	2.77	

** Significant at $p = .01$; *Significant at $p = .05$

126

TABLES OF ANALYSIS OF COVARIANCE

Source	D.f.	Sy^2	Sxy	Sx^2	D.f.	S.S.	M.S.	F
						----Adjusted----		
Table VI								
Concentrations	1	6453.03	0	0	1	6453.03	6453.03	
Racks	19	135.43	0	0	19	135.43	7.13	
Conc. x Racks	19	127.93	0	0	19	127.93	6.73	1.05
"Duplicate"Tubes	40	438.88	-1861.75	1060.00	39	113.73	2.92	2.30*
Table VII								
Concentrations	1	6381.38	0	0	1	6381.38	6381.38	
Racks	19	98.81	0	0	19	98.81	5.20	
Conc. x Racks	19	56.31	0	0	19	56.31	2.96	1
"Duplicate"Tubes	40	1408.88	-3076.75	10660.00	39	520.85	13.36	1
Table VIII								
Concentrations	1	7296.20	0	0	1	7296.20	7296.20	
Racks	19	119.01	0	0	19	119.01	6.26	
Conc. x Racks	19	165.92	0	0	19	165.92	8.73	1
"Duplicate"Tubes	40	705.26	-2175.00	10660.00	39	261.49	6.70	1.30

* Significant at $p = .05$

KEEP IT SIMPLE - - IT PAYS

Axel U. Sternlof
American Steel and Wire Division
United States Steel Corporation

Many of our well established industrial enterprises have learned of the near-fabulous mushroom growth of statistical quality control as applied to manufacturing processes. They have seen stories in the various trade journals of the thousands of dollars that have been salvaged from the industrial junk-heaps as well as of the tremendous improvement that has been made in product quality through the use of this technique. They have heard of what can be accomplished through the use of statistical techniques involving correlation and regression; they have also heard of the fantastic savings through statistical design and analysis of experiments. Consequently, these modern management tools have been looked upon as a veritable panacea for all manufacturing ills and a great many manufacturers have been guilty of putting "the cart before the horse". The application of these so-called advanced techniques of quality control to manufacturing processes unquestionably has its place in our every day industrial life, and intricate problems are solved in a comparatively short time. But what about the basic or the fundamental processes? Are they in a state of statistical control? If so, is that control satisfactory in relation to the tolerances or the specifications? In many cases it is the application of the simple or fundamental techniques of quality control that reveals the gold in the industrial mine.

Let us take a look behind the scenes of an electrical wire and cable plant. Here we find a typical job shop with hundreds of different products, using many different materials, frequent set-up changes, with few product duplications, many of them being manufactured to trade specifications, to customer specifications, etc., which frequently are vastly different.

Then, too, we have the situation where we must judge the whole product by what is evident from an examination of end conditions only. We cannot cut a cable in the middle to check its characteristics. However, in spite of all these variables, there are applications of quality control that can be used to pay off handsomely. Let us look into this situation and see what we can learn.

First, we have the drawing of the copper wire to the tolerances specified by the Product Engineering Department. I have selected one size out of our actual experience in the plant to illustrate how the dividends materialize. This particular wire was drawn to the tolerances of 0.0122" to 0.0123" and had been processed to this specification for years. Average and range charts were kept on the diameter and the physical characteristics such as per cent elongation and tensile strength. In the matter of this presentation we are interested solely in the diameter of the wire. Charts for this particular characteristic showed very good control. Now and then a point on the average or the range chart would go out of control as the drawing die deteriorated.

Shortly after the installation of quality control in the plant, we had an inquiry from the engineers in relation to the diameter of this particular wire. It was explained a customer had complained on a product

in which this wire size was used. The Quality Control Organization was requested to look into the matter and after a detailed survey which took into consideration every variable that entered into the process, we concentrated on the wire diameter. Our original data sheets were available for all characteristics for several hundred tests and after exoneration of all the other variables we went back to study the diameter of the stranded copper conductor. On investigation we found that the specification called for a seven-wire stranded conductor with a nominal diameter of 0.0360" and tolerances of 0.0351" to 0.0369". In a seven-wire concentric stranded conductor, that simply meant that each wire should have a nominal diameter of 0.0120". When the data was charted for the stranded conductor diameter the interesting situation as shown in Figure #1 was portrayed. This chart represents 100 reels of stranded wire, the individual measurements being arranged in groups of 4 for control chart purposes. Figure #2 shows our quality report to the engineers which gives a very good comparison of product versus specifications. The high "per cent above specifications" does not mean bad product. It does show, however, that since the product is sold by the foot, we were giving away copper. Continuing our investigation it was disclosed that at some time in the distant past the tolerances of 0.0122" to 0.0123" had been stipulated as necessary in order to produce a finished wire with the nominal of 0.0120". The extra 0.0002" approximately were added to take care of "pull down" in the subsequent processes of tinning and stranding. It was plainly evident from hundreds of measurements made by the Quality Control observers that this allowance was not necessary. New tolerances of 0.0119" to 0.0121" were finally adopted, which meant a saving of over 4% in copper. This saving was very easily translated into dollars and cents for a report to management. This was merely one of many similar illustrations of factual data proving conclusively that opinion can be wrong.

To complete this picture let us look at Figures #3 and #4 which show the results of the process after the change had been put into effect. Figure #5 is a chart showing one-half of "before and after" charts for comparative purposes.

For the second half of this presentation let us compare two sampling plans; the first is illustrative of a rather complex plan; the second, of a simple plan. We will compare the amount of inspection involved under each plan and the end results of both. The product involved is an electrical wire known commercially as "building wire", the insulation being of the type known as thermoplastic and is in the 2/64" insulation thickness class.

The inspection station is located at the automatic coiling machine where a reel of wire of several thousand feet in length is automatically coiled and cut into coils of 500 feet. For the purpose of this comparison each reel is considered a "lot" and the size of the lot is dependent on the number of coils obtained from the reel. As each coil comes from the machine, a piece approximately a foot long is cut from the end of every coil. It is from these end pieces that we are to select our samples at random, measure the characteristic, and determine the fate of the lot under both of these inspection plans. When the sample has been selected under the first plan and the individual pieces have been tested, all the pieces are then cut to the same length and the sample selected at random for the second plan and the pieces tested again. In this manner there will be no bias or prejudice in the selection of the various

samples.

For this comparison, the lots are grouped into sets of 25 reels each and we have 12 such lots representing a grand total of 4,954 coils. Furthermore, the bare conductor in these lots varied in size and type, such as solid and stranded. The common criteria here are type and thickness of insulation.

The plan which I refer to as being rather complex is under consideration by a committee from the I.P.C.E.A. (Insulated Power Cable Engineer's Association) and is styled as Project #356. The following is quoted from the plan:

"Section 2, Minimum Wall as Criterion
"......... The normal sigma for a manufacturer entering this plan shall be the 100% column.
-. . . . .

"Section 4, Sampling Plan
"See Tables I and II

"Instructions For Using Variables Sampling Plan

"For given type of insulation and lot size, take sample of size of first value given in cumulative sample size column. If the average of the values in this sample is equal to or above the corresponding lowest acceptable average given in the proper sigma column, the lot is accepted. If the average is below this value, a second sample is taken such that the total sample size is equal to the second value given in the sample size column. If the average of the total sample is equal to or above the corresponding lowest acceptable average, the lot is accepted. If the average is below this value, the lot is rejected.

"Instructions for Changing Sampling Plans

"If the ranges of four consecutive lots exceed the corresponding R_t value, the plans for the next higher sigma must be used.

"If the ranges of twenty consecutive lots are all equal to or less than the corresponding R_T value, the plans for the next lower sigma may be used..............."

Table I (In Part) (1)
Variables Sampling Plan

Insulation	Lot Size	Cumulative Sample Size	Lowest Acceptable Average In Mils				Absolute Minimum Mils
			Sigma Equal To, In Percent				
			50	75	100	125	
Thermoplastic 31 Mils In Thickness	1 - 3*	2	27.0	28.3	29.5	30.8	
		4	26.7	27.8	28.9	30.0	
	4 - 10	2	27.0	28.3	29.5	30.8	23.0
		4	26.8	28.0	29.1	30.3	
	11 and Over	5	27.1	28.4	29.7	31.0	
		10	27.0	28.2	29.4	30.6	

*Each coil is tested separately taking one (or two) pieces from each end of coil.

Table II (1)
Requirements for Changing Sampling Plan

Insulation	Lot Size	Cumulative Sample Size	Sigma Equal To, In Percent							
			50		75		100		125	
			R_t	R_r	R_t	R_r	R_t	R_r	R_t	R_r
Thermoplastic 31 Mils In Thickness	1 - 3*	2	1	-	2	3	3	4	-	6
		4	3	-	4	4	5	6	-	8
	4 - 10	2	1	-	2	3	3	4	-	6
		4	3	-	4	4	5	6	-	8
	11 and Over	5	3	-	4	4	6	6	-	8
		10	4	-	5	5	7	7	-	10

*Each coil is tested separately taking one (or two) pieces from each end of coil.

So much for the details of the I.P.C.E.A. Project #356 plan. The so-called simple plan is akin to that recommended by the ASTM Task Force in their tentative specifications for the control of copper wire physical properties such as per cent elongation and tensile strength. This plan suggests that, regardless of the lot size, a random sample of four is selected for the measurement of the characteristic involved. I feel that it is imperative to quote from the notes in the A.S.T.M. specifications for "assurance" that the plan will work; "Note 5. Specifications for Hard Drawn Copper Wire B 1 - 53T.

"Cumulative results secured on the product of a single manufacturer indicating continued conformance to the criteria, are necessary to insure an over-all product meeting the requirements of these specifications. The sample sizes and conformance criteria given for the various characteristics are applicable only to lots produced under these condi-

tions" (Ref. 2). The intention from the quote is quite plain; namely, that a control chart for average and range faithfully kept is a big asset leading toward quality assurance. The A.S.T.M. plan states that a second sample be taken under certain conditions. However, even though a second sample may be required at times, the conformance criteria remain the same as far as the characteristics are concerned.

In order to conserve both time and space, the details of the many data sheets are not reproduced here, a summary being used in its stead for the comparison of the two plans under consideration. Figures #6 and #7 might be typical of the sort of control chart that might be used under the I.P.C.E.A. plan. The solid lines shown on the charts are not control limits; they are the criteria limits that are determined by the sigma column and sample size that might be in use at the moment.

Let us refer to Figure #6. This shows pictorially the data of the first 25 lots of reels inspected. Note in particular that the range of the first 20 lots in succession is better than that called for under the 100% sigma column, and note, also, that two lots stand rejected. Consequently, at lot #21 we are entitled to inspect under the criteria found in the 75% sigma column, which criteria will now accept the two lots that were rejected under the 100% sigma column. Figure #7 shows the performance of the second 25 lots and note that one lot stands rejected under the 75% sigma column criteria. During the inspection of the third set of 25 lots, we were entitled to go to the 50% sigma column, the lowest of the plan, and now we can pass the lot that was rejected under the 75% sigma column.

As we examine the plan and the charts shown here, I would like to call your attention to the following facts:
1). In quality control work, it is quite the customary practice that once a process has established itself, control charts, with limits as previously determined or set by the process itself, are placed into operation in the plant. Under the IPCEA Project #356 no such limits are possible: only specification minimum can be shown and these can be located on the chart only after the actual testing. The reason for this is that the minimum specification can be determined only by the final sample size and the particular sigma column being used at the time. HERE IT IS THE PLAN, NOT THE PROCESS, THAT DETERMINES THE LIMIT.

2). The criteria of both the minimum wall thickness and the range are not fixed, but are determined by the sigma column and the sample size used at the moment.

3). It is my understanding (from a member of the Project Task Force) that this plan is not intended to be used as a control for manufacturing processes but, rather, as a plan for the inspection of incoming material.

If this be the case, I can imagine a vigorous protest from those manufacturers whose processes are controlled at the criteria of the 50% sigma column, only to see that very same product rejected by the customer who inspects under the standards found in the 100% sigma column. This plan then penalizes the manufacturer who has learned to control the process with very low variation in the minimum wall thickness. I am not contesting the statistical soundness or validity of the plan for it is

133

not the purpose of this paper to enter into the arena of academic argument. I do, however, seriously question its _practicability_ from an industrial viewpoint and object strenuously to its required use because the plan would increase inspection costs considerably. Such increase must be passed on to the customer in the price he pays for the product.

Let us now compare the results under the two plans. I wish to point out that the measurements were made by using a pin gauge and recording these to the nearest .0001". Averages were carried out to six decimal places and no "rounding off" took place until this final summary. Three hundred lots or 4,954 coils were sampled under each plan. The time period covered by the data for this comparison is not confined to a week's run, but was purposely spread out over several months. We must realize that the "over-all" distribution will best be approximated if the data under consideration covers many operators, machines,set-ups, etc. Today's run may be at the lower portion of the over-all distribution curve and next week we may be operating at the upper portion. To estimate the standard deviation from the data obtained from any one set-up or run and translate this information as being typical of the process over a length of time might lead to false conclusions.

Plan	Coils Tested	$\bar{X}$	σ'	$\bar{X} - 3\sigma'$	$\bar{X} + 3\sigma'$
IPCEA	1881	.0297	.001057	.0265	.0329
n = 4	1220	.0300	.001036	.0269	.0331

In order to calculate the standard deviation overall, it was necessary to revamp the data under the IPCEA plan into sample sizes of 2, 4, 5 and 10 pieces, make the calculation for each sample size and then weight the totals for the final results. This standard deviation overall (σ_0) calculation is the same as that used by the A.S.T.M. Task Force and takes into consideration the following:

1). The standard deviation of the averages .... $\sigma_{\bar{x}}$

2). The standard deviation within the groups .... σ_w

3). A correction factor (n = sample size) .... $\sigma_{w/n}$

4). The standard deviation among the groups .... $\sigma_{\bar{x}b}$

5). The standard deviation over all .... σ_0

The formula for this calculation is as follows:

$$\sqrt{\left(\sigma_{\bar{x}}^2 - \frac{\sigma_w^2}{n} \right) + \sigma_w^2} = \sigma_0$$

Applying the above formula and calculating the standard deviation overall for the data of the two plans, we have:

Plan	Coils Tested	$\bar{X}$	σ_0	$\bar{X} - 3\sigma_0$	$\bar{X} + 3\sigma_0$
IPCEA	1881	.0297	.001228	.0260	.0334
n = 4	1220	.0300	.001449	.0256	.0343

134

A word of further explanation might now be in order as we compare the above results. It seems evident that the "100% sigma" column of the IPCEA plan is the average obtained under the industry-wide survey conducted by the Project #356 committee. The "absolute minimum" of 23 mils stated in the plan could be the industry-wide average minimum wall thickness minus three standard deviations. In quality control language we say that this value could occur about three times in one thousand. We must remember that this survey was made from data submitted by many producers testing their own product which was manufactured to meet the present trade specification of .028" or 28 mils as a minimum wall thickness.

From the above tabulation, we can see that the results obtained under both of the sampling plans are practically identical. The only significant difference between the two plans is indicated in the relative amounts of inspection necessary to obtain these identical results; the IPCEA Project #356 required approximately 50% more Testing than did the ASTM n = 4.

So from this comparison it is quite evident to see that to

"Keep It Simple - It Pays".

References:

      (1)   IPCEA Project #356
             Second Draft April 14, 1955
             "Dimensional Tolerances".

      (2)   ASTM Standards on Metallic Electrical Conductors
             October 1954, page 33, Note 5.

Figure #1   Strand Diameters
n = 4

136
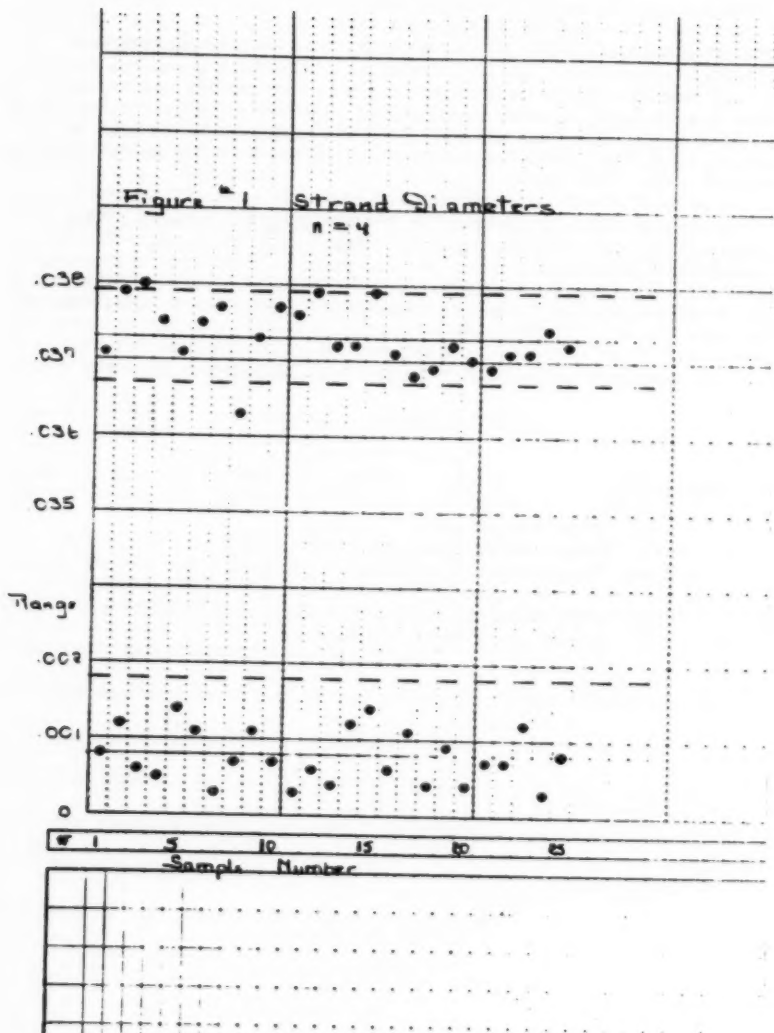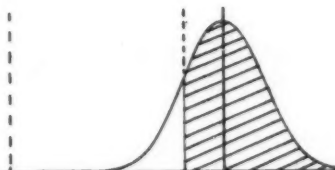
% Below Spec. Min.	0
% Above Spec. Max.	84.85
% Within Specs	15.15

Process Center   .0373"
Process Capabilities .0361" - .0385"

LSL - Lower Specification Limit=   .0351"
USL - Upper Specification Limit=   .0369"
$\overline{X}'$ - Mean of Specifications=   .0360"
$\overline{X}$ - Mean of Observations=   .0373"
σ' - Standard Deviation= .000388
X1 - $\overline{X}$ - LSL=   .0373 - .0351 = .0022
X2 - USL - $\overline{X}$=   .0369 - .0373 = .0004
C - $\overline{X}$ - $\overline{X}$-   .0373 - .0360 = .0013   Amount production is
   off-center relative to mean of specifications.

                                                History
$$\frac{X1}{\sigma'} = \frac{.0022}{.000388} = 5.67$$ Year   Average   Std. Dev.   % B. S.

$$\frac{X2}{\sigma'} = \frac{.0004}{.000388} = -1.03$$

$$\frac{C}{\sigma} = \underline{\hspace{2cm}} =$$

137

Figure #3 Strand Diameters
n = 4

138
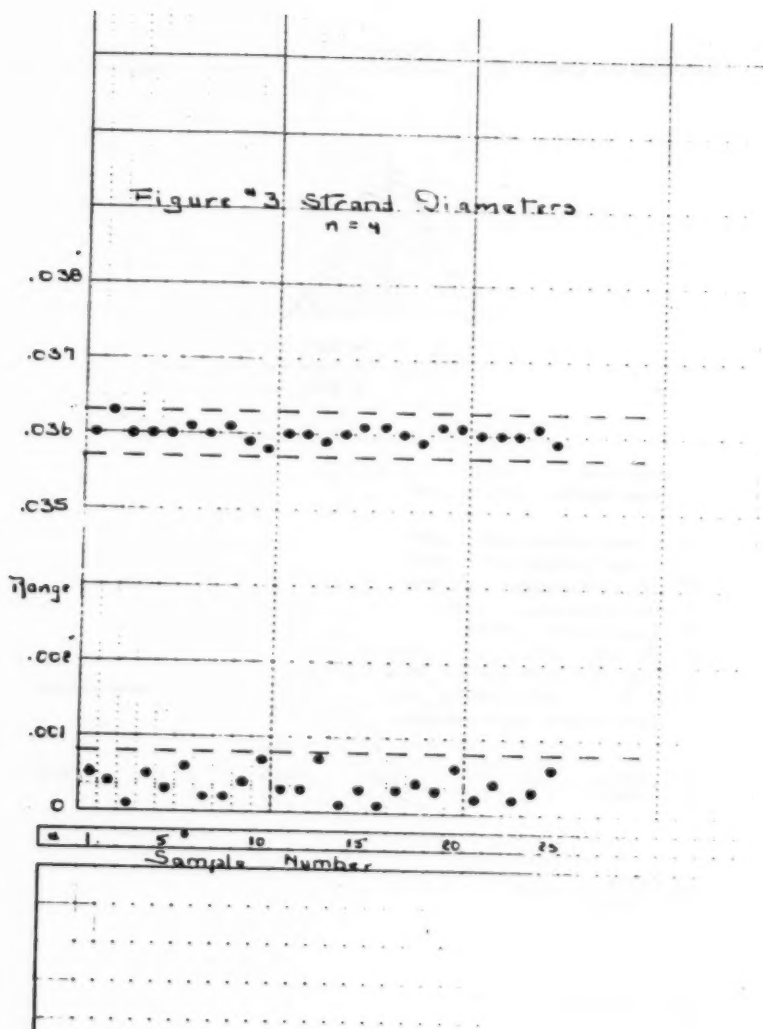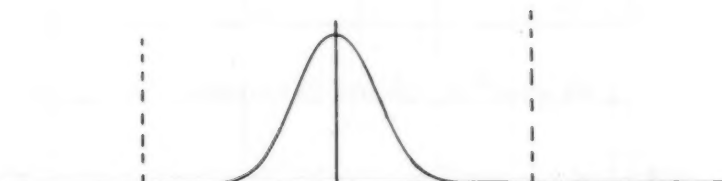
% Below Spec. Min.   0
% Above Spec. Max.   0
% Within Specs.   100

Process Center   .0360"
Process Capabilities .0355" - .0365"

LSL - Lower Specification Limit- .0351"
USL - Upper Specification Limit- .0369"
$\overline{X}'$ - Mean of Specifications-   .0360"
$\overline{X}$ - Mean of Observations-   .0360"
σ'- Standard Deviation- .000175
$X1 = \overline{X} - LSL$=   .0360 - .0351 = .0009
$X2 = USL - \overline{X}$=   .0369 - .0360 = .0009
$C = \overline{X} - \overline{X}$=   .0360 - .0360 =   0   Amount production is
off-center relative to mean of specifications.

                                    History
$\dfrac{X1}{\sigma'} = \dfrac{.0009}{.000175} = 5.14$ Year   Average   Std. Dev.   % B. S.

$\dfrac{X2}{\sigma'} = \dfrac{.0009}{.000175} = 5.14$

$\dfrac{C}{\sigma} = \underline{\quad\quad} =$

139

Figure 5 Strand Diameters
n = 4

Figure 7-4   IPCEA 5-96

141

Figure # 7   I.P.C.E.A. # 35L

142
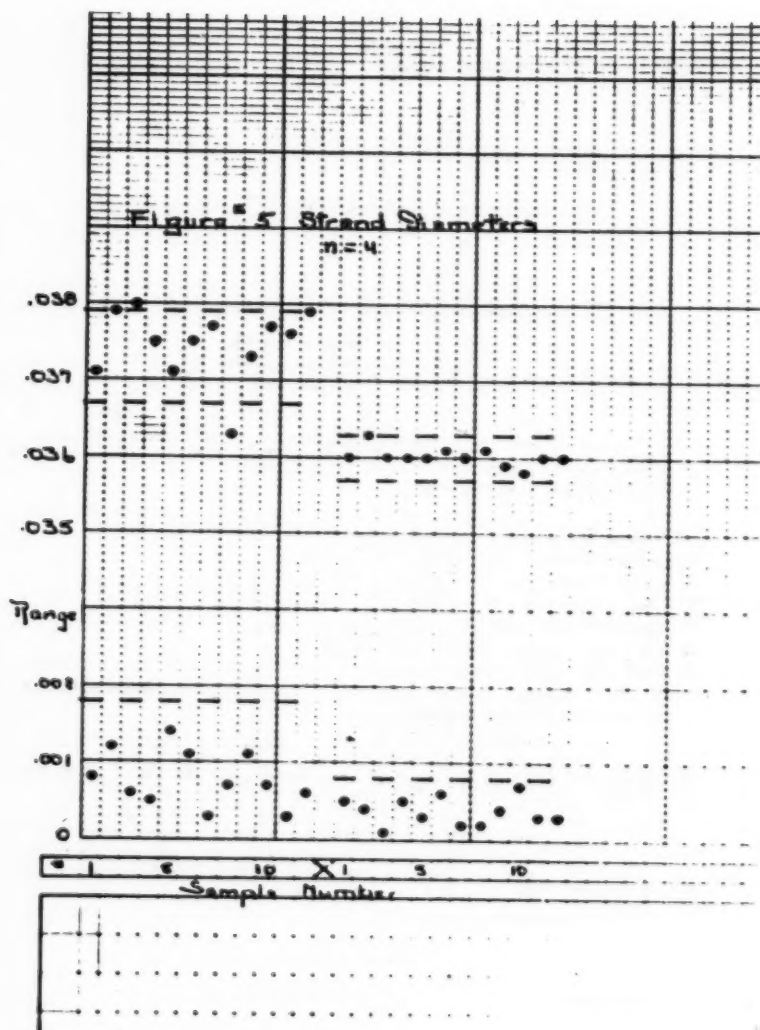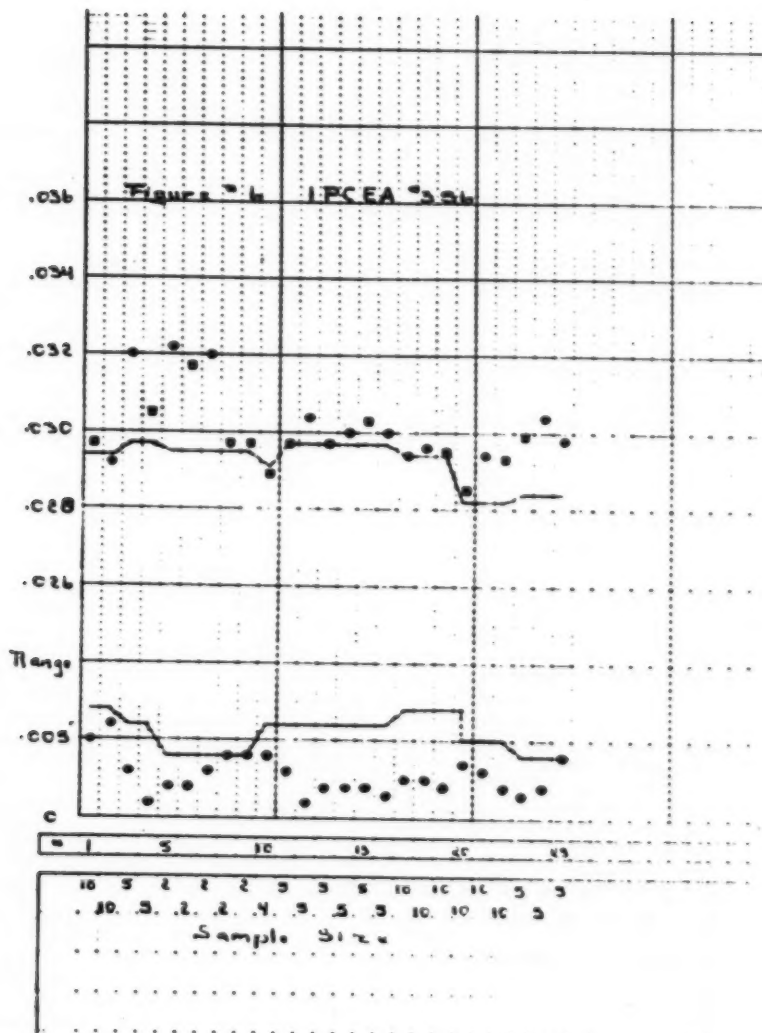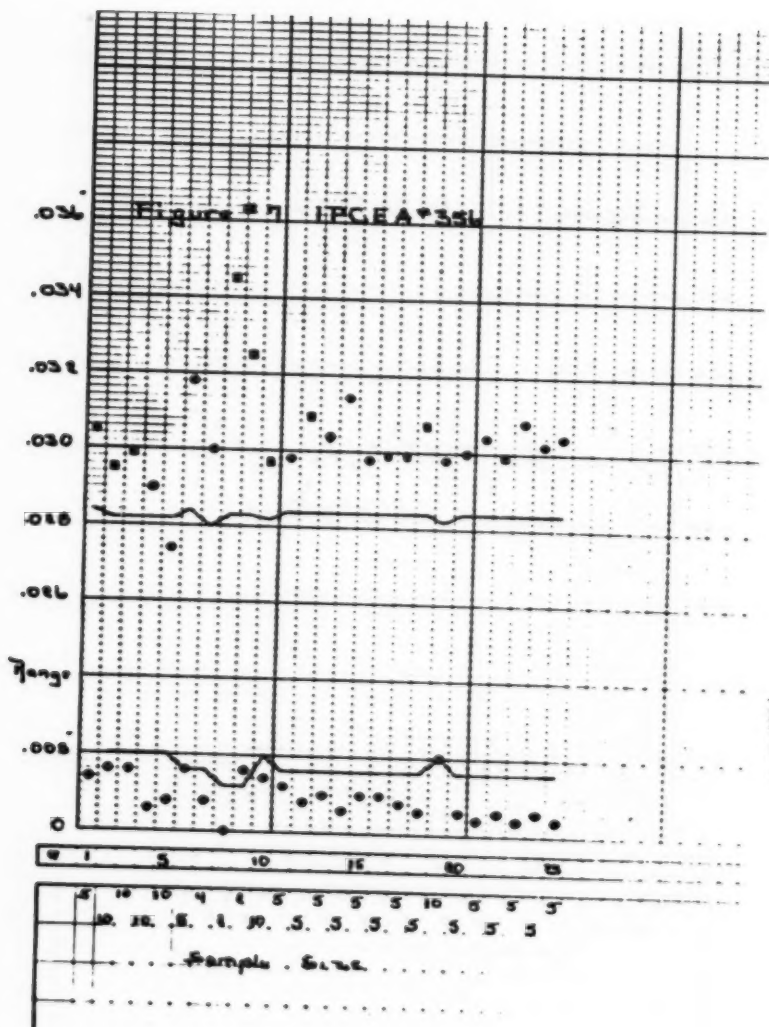
RECENT RESEARCH ON STATISTICAL PROBLEMS IN SUBJECTIVE TESTING

Ralph Allan Bradley
Virginia Polytechnic Institute

1. Introduction

There is widespread interest in the design, conduct, and analysis of experiments involving the subjective opinions of samples and panels of individuals. Applications arise in food processing, photography, distilling and brewing, textile research, wood technology, petroleum products research, and in a host of other areas of research.

Many statistical problems occur in subjective testing that require further attention. We shall present here the results of recent research, with illustrative examples, on techniques that deal with the sensitivities of scoring scales and with the variabilities of judges or respondents using scoring scales. While these methods relate to the use of scoring scales as they are widely employed, we also remind the reader that it may often be desirable to employ simple ranking methods in subjective testing.

2. Sensitivity Comparisons

In the development of scoring scales and other experimental techniques, it is often desirable that two alternative methods be compared. Cochran [1] discussed the comparison of different scales of measurement for experimental results and indicated where further research was required. We have provided means of comparing the sensitivities of similar experiments in two recent papers (Schumann and Bradley [2], Bradley and Schumann [3]). This recent research permits a test on the equality of the parameters of non-centrality of F-distributions associated with tests of treatment equality in two independent but parallel experiments containing the same set of treatments in identical experimental designs. The experiments may differ in the scoring scale used or in some other criterion of measurement that does not interact with treatments.

Good experimental data to illustrate the method to be presented in this section were published recently by Kauman, Gottstein, and Lantican [4] who were interested in the quality evaluation of dried veneer. They used two evaluation schemes: "numerical" with a zero-score being excellent and a score of fifty being very bad and based on assigned scores for various types of degrade, and "subjective" with quality appraisal ratings from zero, excellent, to eight, very bad. Twenty selected sheets of veneer were evaluated by three observers, twice with each scheme with repeat observations spaced by several days and orders of presentation of sheets changed. Complete tables of scores are given in the reference; we require the summaries of Table 1 below.

Table 1
Analyses of Variance for Quality Ratings*

Factor	Degrees of freedom	Numerical scheme		Subjective scheme	
		Sum of squares	Mean square	Sum of squares	Mean square
Sheets (S)	19	12826.16	675.1	336.90	17.73
Observers (O)	2	170.72	85.36	3.70	1.852
Repetitions (R)	3	168.13	56.04	0.61	0.2042
Interaction(SO)	38	823.61	21.67	30.12	0.7928
Error (SR)	57	595.37	10.45	22.13	0.3884

* A reproduction of part of Table 6, Kauman, Gottstein, and Lantican [4], page 148.

We apply our method to compare the sensitivities of the two evaluation schemes, the test being one to ascertain which of the two schemes is more sensitive for the exhibition of quality differences among the sheets of veneer. Detailed discussions of the procedure used are given in the cited references, and we here illustrate the procedure only. We restrict our interest to the twenty sheets of veneer in the experiment and assume Model I of the analysis of variance.

The expected value of the mean square for sheets is

$$(1.2) \qquad E[M.S.(S)] = \sigma^2 + k \sum_{i=1}^{t} \tau_i^2/(t-1)$$

in general for t sheets and k observations on each sheet. τ_i is the "effect" of sheet i, i = 1,...,t. In the examples,

$$(2.2) \qquad E[M.S.(S)] = \sigma^2 + 6 \sum_{i=1}^{20} \tau_i^2/19.$$

σ^2 is the expectation of the error mean square in both (1.2) and (2.2). The parameter of non-centrality of the F-test for sheets is, in general,

$$(3.2) \qquad \lambda = k \sum_{i=1}^{t} \tau_i^2/2\sigma^2$$

and, in the examples,

$$(4.2) \qquad \lambda = 6 \sum_{i=1}^{20} \tau_i^2 / 2\sigma^2$$

when the F-density is written

$$(5.2) \qquad f(F) = (a/b)^a [B(a,b)]^{-1} e^{-\lambda} F^{a-1} (1+aF/b)^{-(a+b)}$$
$$\cdot \, _1F_1[a+b, \; a, \; a\lambda F/b(1+aF/b)], \; 0 \leq F \leq \infty$$

where F has 2a and 2b degrees of freedom, $_1F_1$ is the confluent hypergeometric series, and B represents the beta function. λ is a parameter expressing the magnitudes of treatment effects in a scale in terms of the experimental error associated with the scale. λ is the appropriate parameter to measure the sensitivity of a scale.

We shall test the hypothesis, H_0: $\lambda_1 = \lambda_2$, against the alternative, H_a: $\lambda_1 \neq \lambda_2$, using the subscript 1, for the numerical scheme, and 2, for the subjective scheme.

To apply the test, we compute the two F-ratios with 19 and 57 degrees of freedom (Now a = 9.5, b = 28.5.) and obtain

$$F_1 = 64.60 \qquad \text{and} \qquad F_2 = 45.65.$$

The statistic used is

$$(6.2) \qquad w = F_1/F_2 = 64.60/45.65 = 1.42.$$

The distribution of w under H_0 depends on $\lambda = \lambda_1 = \lambda_2$ which in general is unknown. In practice it is clear that the test is not very sensitive to small changes in λ and we, in fact, estimate λ from the data using

$$(7.2) \qquad \hat{\lambda}_i = a(F_i - 1), \; i = 1, \; 2.$$

In the examples,

$$\hat{\lambda}_1 = 9.5(64.60-1) = 604.2$$

and

$$\hat{\lambda}_2 = 9.5(45.65-1) = 424.2.$$

We take $\hat{\lambda}$ to be the average of $\hat{\lambda}_1$ and $\hat{\lambda}_2$,

$$(8.2) \qquad \hat{\lambda} = \tfrac{1}{2}(604.2 + 424.2) = 514.2.$$

145

A table of values w_0 such that $P(w > w_0 | H_0) = .05$ is given by Bradley and Schumann in the cited references. To enter this table, one requires

$$(9.2) \qquad a' = (a+\lambda)^2/(a+2\lambda) = (9.5 + 514.2)^2/(9.5 + 1028.4)$$

$$= 264.2$$

and $b = 28.5$. The table is symmetric in the sense that $w_0(a',b) = w_0(b,a')$, and we obtain $w_0 \approx 1.85$ by consulting the table. Now H_a, as postulated, is two-sided and hence the significance level being used is .10. w in (6.2) does not exceed w_0, and consequently we do not reject H_0 at the 10 per cent level of significance. We are in accord with the authors (Kauman et al.) who state "the present experiment has shown that the subjective evaluation can yield results of an accuracy approaching that of the numerical scheme, although the accuracy of the latter was slightly superior."

Applications are limited since a table is only available for a one-sided 5 per cent level test. Schumann is preparing additional tables that are nearly complete.

The procedure may also be applied to compare multiple correlation coefficients under certain conditions and an example is given in reference [3].

3. Judge Variability and Judge Comparisons

When items are scored in subjective experimentation, there is no knowledge of the "true worth" of the sample in the units of the scoring scale. It is then difficult to assess the judging ability of a judge. Russell and Bradley [5] have provided means of estimating the variability of a judge in terms of the deviations of his scores for an item from those of the remaining judges but permitting a judge a possible constant bias in his assignment of scores. Similar procedures were considered by Grubbs [6] and Ehrenberg [7].

Consider a two-way classification with t items or treatments and r judges. The model with fixed effects is

$$(1.3) \qquad y_{ij} = \mu + \tau_i + \beta_j + \epsilon_{ij}, \ i = 1,\ldots,t, \ j = 1,\ldots,r$$

where y_{ij} is the score assigned by the jth judge to the ith item, μ is the grand mean, the average level of judging, τ_i is the effect of the ith item, β_j is the effect (or bias) of the jth judge, and ϵ_{ij} are independent normal variates with zero means. Contrary to the usual model of analysis of variance, we admit the possibility of heterogeneous error variances in the sense that

146

$$(2.3) \qquad E(\epsilon_{ij}^2) = \sigma_j^2,$$

σ_j^2 is the variance of the jth judge and is to be estimated.

The estimator of σ_j^2 to be used is

$$(3.3) \qquad \hat{\sigma}_j^2 = \frac{rG_j}{(t-1)(r-2)} - \frac{E}{(t-1)(r-1)(r-2)}$$

where

$$(4.3) \qquad G_j = \sum_{i=1}^{t} (y_{ij} - y_{i.} - y_{.j} + y_{..})^2$$

and

$$(5.3) \qquad E = \sum_{i=1}^{t} \sum_{j=1}^{r} (y_{ij} - y_{i.} - y_{.j} + y_{..})^2,$$

the latter being the error sum of squares from the analysis of variance of the two-way classification. $\hat{\sigma}_j^2$ is an unbiased estimator of σ_j^2 but, like an estimate of a variance component, may occasionally be negative. In (4.3) and (5.3), $y_{i.}$ is the average of scores for treatment i, $y_{.j}$ is the average of scores assigned by the jth judge, and $y_{..}$ is the average of all scores. The requirement that ϵ_{ij} in (1.3) be normal is only met approximately in use of a discrete scoring scale but does not affect the estimation of σ_j^2. In later paragraphs of this section, we shall assume that departures from normality do not seriously affect our test procedures.

We shall again illustrate this work using the data of Kauman et al. The detailed example is for Test 1 using the subjective scheme. Scores are listed in Table 2. In Table 3 we show values of the residuals, $(y_{ij} - y_{i.} - y_{.j} + y_{..})$. Values of G_j and E are given in the lower margin of Table 3 and are obtained by accumulating the squares of entries in the columns above as required in view of (4.3). $E = \sum_{j=1}^{r} G_j$ and was so obtained. The values of $\hat{\sigma}_j^2$ computed using (3.3) are listed in Table 4 along with those for the other three tests of Kauman et al. To illustrate the computations, we use observer A and obtain

$$\hat{\sigma}_1^2 = \frac{3(8.36)}{(19)(1)} - \frac{22.50}{(19)(2)(1)} = 0.72.$$

A test of homogeneity of variances is possible only when $r = 3$. The only situation wherein the estimators $\hat{\sigma}_j^2$ of σ_j^2 are maximum likelihood estimators is when $r = 3$ and then an approximate test may be made. Consider the hypothesis,

Table 2

Quality Ratings for the Subjective Quality Evaluation, Test 1*

Sheet No.	Observers A	B	C	$y_i.$
1	3	3	3	3.00
2	7	5	7	6.33
3	6	5	5	5.33
4	7	8	7	7.33
5	1	2	3	2.00
6	5	5	5	5.00
7	5	6	5	5.33
8	3	5	4	4.00
9	4.5	5	5	4.83
10	6	7	7	6.67
11	5	4	4	4.33
12	8	7	8	7.67
13	5	7	5	5.67
14	1	2	2	1.67
15	7	7	7	7.00
16	1	3	4	2.67
17	4	3	3	3.33
18	6	6	5	5.67
19	5	5	7	5.67
20	3	3	2	2.67
$y.j$	4.62	4.90	4.90	$y..$ 4.81

Table 3

Values of $(y_{ij}-y_i.-y.j+y..)$ for the Subjective Quality Evaluation, Test 1

Sheet No.	Observers A	B	C
1	.19	-.09	.09
2	.86	-1.42	.58
3	.86	-.42	-.42
4	-.14	.58	-.42
5	-.81	-.09	.91
6	.19	-.09	-.09
7	-.14	.58	-.42
8	-.81	.91	-.09
9	-.14	.08	.08
10	-.48	.24	.24
11	.86	-.42	-.42
12	.52	-.76	.24
13	-.48	1.24	-.76
14	-.48	.24	.24
15	.19	-.09	-.09
16	-1.48	.24	1.24
17	.86	-.42	-.42
18	.52	.24	-.76
19	-.48	-.76	1.24
20	.52	.24	-.76
G_j	8.36	7.07	7.07

E=22.50

* From Table 3, Kauman, Gottstein, and Lantican [4], page 135.

148

Table 4

Estimates of Variance and X_2^2 to Test for Homogeneity of
Observer Variances for All Four Tests of Kauman et al.

Tests	Observer Variances, $\hat{\partial}_j^2$			Error Mean Square, $\hat{\partial}^2$	χ^2
	A	B	C		
Subjective Test 1	.72	.53	.53	.59	.14
Subjective Test 2	.46	.58	.74	.59	.26
Numerical Test 1	4.58	10.79	33.02	16.13	6.42*
Numerical Test 2	2.19	26.40	21.69	16.84	4.14

$$H_o: \sigma_1^2 = \sigma_2^2 = \sigma_3^2,$$

and the alternative,

$$H_a: \sigma_j^2 \neq \sigma_k^2 \text{ for some } j \text{ and } k, \; j, \; k = 1, \; 2, \; 3.$$

The likelihood ratio test statistic, distributed approximately as a X^2-variate with 2 degrees of freedom for large samples, is

$$(6.3) \quad X^2 = -(2.3026)(t-1)[2 \log(t-1) + \log(\hat{\partial}_1^2\hat{\partial}_2^2 + \hat{\partial}_1^2\hat{\partial}_3^2 + \hat{\partial}_2^2\hat{\partial}_3^2)$$

$$- 2 \log E + \log 4/3]$$

$$= -(2.3026)(19)[2 \log 19 + \log\{(.72)(.53)$$

$$+(.72)(.53) + (.53)(.53)\} - 2 \log 22.50 + \log 4/3]$$

$$= 0.14.$$

The multiplier, 2.3026, in (6.3) is included so that common logarithms may be used in the computation of X^2. The small value of X^2 indicates that the observers may be taken to have homogeneous variances.

In Table 4, we have included values of X_2^2 for all four tests and show also values of $\hat{\partial}^2$, the error mean square from the analysis of variance. Note that only in one of the numerical tests was X_2^2 significant at the 5 per cent level of significance. The estimates of variance in the numerical

149

scheme are considerably larger than in the subjective scheme. This does not of course suggest a preference for the subjective scheme but is merely a result of the scales used in the scoring methods. The appropriate method of comparing the scales is the one given in the preceding section.

Observer A was the only observer with previous experience in judging veneer except for brief training sessions before the experiment began. Another test, and this is an exact test, is possible. Consider the null hypothesis

$$H_o: \sigma_1^2 = \sigma^2, \text{ given } \sigma_2^2 = \ldots = \sigma_r^2 = \sigma^2,$$

and the alternative,

$$H_a: \sigma_1^2 < \sigma^2, \text{ given } \sigma_2^2 = \ldots = \sigma_r^2 = \sigma^2.$$

the statistic used is

$$(7.3) \qquad F = \frac{r(r-2)G_1}{(r-1)E - rG_1}$$

with $(t-1)$ and $(t-1)(r-2)$ degrees of freedom. H_a may have either of the possible one-sided forms or be two-sided. For the form of H_a shown, small values of F are significant. In the example, there is no point in testing H_o versus H_a in this test in view of the homogeneity of variances demonstrated above. However, we shall proceed in order to illustrate the method.

$$F = \frac{(3)(1)(8.36)}{(2)(22.50) - 3(8.36)} = 1.26.$$

With 5 per cent level of significance, we obtain the critical value of F by reading the tabular value of F with $(t-1)(r-2) = 19$ and $(t-1) = 19$ degrees of freedom from an F-table and taking its reciprocal. In this case, values of $F \leq 1/2.13 = .47$ are significant at significance level .05. This F-test is useful, for example, in considering the addition of a new judge to an established taste panel. Values of F with their upper and lower bounds for significance may be plotted to yield a rough control-type of study on the performance of a given judge in successive experiments. The fact that other judges do not necessarily have homogeneous variances should not seriously affect the procedure. A form for F, alternate to (7.3), may sometimes be preferred. This form is

$$(8.3) \qquad F = \frac{E + (t-1)(r-1)(r-2)\hat{\sigma}_1^2}{rE - (t-1)(r-1)\hat{\sigma}_1^2}.$$

150

4. Discussion and Summary

We have discussed two procedures that are of use in subjective testing wherein scoring scales are used. One deals with the establishment of suitable scoring procedures and experimental techniques. The other permits an evaluation of the performances of judges using scoring scales and permits plotting of charts showing judges' performances over series of experiments. We now note some additional procedures useful in subjective testing.

Often in subjective experimentation, the items to be compared result or should result from factorial treatment combinations. But subjective evaluations may require the use of incomplete block designs. Kramer and Bradley [8, 9] have provided means of using factorial treatment combinations in incomplete block designs and this may often be useful in taste testing and related areas. An application of these methods is included in these proceedings (H. M. Hill and D. Wheeler [10]).

Ranking methods are often used in sensory testing and may provide very efficient means of experimentation. Paired comparisons are used most frequently and we cite references [11, 12, 13, 14, 15]. Pendergrass [16] in work still unpublished has extended the models for paired comparisons to use in triple comparisons. Some gains in efficiency result in certain situations and his procedures also merit consideration.

References

[1] Cochran, W. G. The Comparison of Different Scales of Measurement for Experimental Results. Annals of Mathematical Statistics 14, 205-216, 1943.

[2] Schumann, D. E. W. and Bradley, R. A. The Comparison of the Sensitivities of Similar Experiments: Theory. Annals of Mathematical Statistics 28, 902-920, 1957.

[3] Bradley, R. A. and Schumann, D. E. W. The Comparison of the Sensitivities of Similar Experiments: Applications. Biometrics 13, 496-510, 1957.

[4] Kauman, W. E., Gottstein, J. W. and Lantican, D. Quality Evaluation by Numerical and Subjective Methods with Application to Dried Veneer. Biometrics 12, 127-153, 1956.

[5] Russell, T. S. and Bradley, R. A. One-way Variances in a Two-way Classification. To be published in Biometrika 45, June 1958.

[6] Grubbs, F. E. On Estimating Precision of Measuring Instruments and Product Variability. Journal of the American Statistical Association 43, 243-264, 1948.

[7] Ehrenberg, A. S. C.   The Unbiased Estimation of Hetero-
geneous Error Variances.   Biometrika 37, 347-357, 1950.

[8] Kramer, C. Y. and Bradley, R. A.   Intra-block Analysis
for Factorials in Two-associate Class Group Divisible
Designs.   Annals of Mathematical Statistics 28, 349-361,
1957.

[9] Kramer, C. Y. and Bradley, R. A.   Examples of Intra-
block Analysis for Factorials in Group Divisible, Par-
tially Balanced, Incomplete Block Designs.   Biometrics
13, 197-224, 1957.

[10] Hill, H. M. and Wheeler, D.   A Designed Experiment to
Evaluate Seven Yarn Lubricants.   Proceedings of the An-
nual Meeting of the American Society for Quality Con-
trol, 1958.

[11] Bradley, R. A. and Terry, M. E.   Rank Analysis of In-
complete Block Designs: I.   The Method of Paired Com-
parisons.   Biometrika 39, 324-345, 1952.

[12] Terry, M. E. and Bradley, R. A.   New Designs and Tech-
niques for Organoleptic Testing.   Food Technology 6,
250-254, 1952.

[13] Bradley, R. A.   Rank Analysis of Incomplete Block De-
signs: II.   Additional Tables for the Method of Paired
Comparisons.   Biometrika 41, 502-537, 1954.

[14] Bradley, R. A.   Rank Analysis of Incomplete Block De-
signs: III.   Some Large-sample Results on Estimation and
Power for a Method of Paired Comparisons.   Biometrika
42, 450-470, 1955.

[15] Dykstra, O.   A Note on the Rank Analysis of Incomplete
Block Designs: Applications beyond the Scope of Exist-
ing Tables.   Biometrics 12, 301-306, 1956.

[16] Pendergrass, R. N.   Ranking in Triple Comparisons.
Ph.D. Dissertation, Virginia Polytechnic Institute Li-
brary, 1957.

QUALITY ASSURANCE TO THE ULTIMATE CONSUMER
IN THE TEXTILE INDUSTRY

J. B. Curley
American Viscose Corporation

The textile industry is a most essential industry to our comfort
and well being. To those not directly concerned with the manufacturing
aspect of this industry, it frequently comes as a surprise that there is
an involved operation that goes into producing the carpet we sink our
feet into, the gracious gown we behold milady wearing or the Bikini
bathing suit we admire.

Speaking of figures, we may take a moment to consider a few. The
carpet we referred to has about 8000 textile fibers in each square inch.
A fiber 400 miles in length is required to make a single square yard of
the dress fabric. In each automobile tire we ride on there may be 4000
miles of textile filaments. The activity of the textile industry con-
sists of arranging and treating these textile fibers such that the final
product has the desired properties of serviceability and aesthetic
appeal.

As can be seen, the quality of the final fabric or rug depends on
the properties of the fibers used, the manner in which these fibers are
put together to form the fabric, and the finishing operations used on the
fabric. Fibers, fabrics, and finishes of widely differing characteristics
are required for a dress or a rug or an automobile tire.

Now the theme of this paper is "how does a manufacturer assure
quality to the ultimate consumer"? I would like to explain to you how a
man-made fiber manufacturer takes steps to assure to an ultimate customer
that the fabric in the dress or the rug is of a functional quality from
which reasonable service may be expected.

There are two programs involved. One is an internal one to see that
a fiber of specific characteristics is manufactured and to control the
manufacture in order that a uniform material is produced. The other pro-
gram is in some respects a novel one. A manufacturer of an "intermediate
product in this vertical textile operation expends effort to see that the
ultimate consumer receives a satisfactory product.

In the abstract, a material can be described by its physical and
chemical properties alone, without reference to what it is made from, how
it was made, or how it will perform. In a practical sense, we rely
rather heavily on the knowledge of how it was made and how it will per-
form.

Let me explain how this applies to our industry. We purchase raw
material against specifications. In addition, we purchase the suppliers
knowledge of the material. This knowledge includes his description and
also how it will perform, that is, his recommended processing behavior.
As a regular check, the supplier sends us a monthly report on certain
properties of the material he has sent to us. For wood pulp this would
be percent cellulose, and percent of all other important components.

Even if the purchaser has complete testing and inspection facilities,
the manufacturer invariably knows more about a product than can be

153

learned by sampling inspections and testing. Some of the many reasons that this situation exists is the frequent inability of the purchaser to sample the material as adequately as the manufacturer. Furthermore, as discussed previously, tests that can completely describe the necessary properties of a material are either time consuming or difficult if not nearly impossible to make. It is for this reason that we use the manufacturers' knowledge of the material which includes the properties known to him from his manufacturing information and evaluations, and his knowledge of how the material can be expected to perform.

On all raw materials we make check tests and inspections but these checks are only to verify the information with which we have been supplied. It requires little logic to see that if material from a particular supplier had a history of meeting the specifications, performed satisfactorily, and was properly described by the reports of the supplier that we would make only occasional tests and inspections on the incoming material. On the other hand, if we were not supplied routine reports of the suppliers tests to determine the extent to which he is meeting the specifications, then we have to carry out a more extensive test and inspection program.

Now to move into the manufacturing operation. Many tests and inspections are made throughout each step of manufacture. With a few exceptions, it can be said that all of these tests and inspections increase our knowledge of the final product. These tests and inspections enable us to control the manufacturing specification which will produce a product of specific properties. This is really the meaning of the idiom that properties of a material cannot be inspected into the material, they must be manufactured into it.

It is in fully utilising these built-in properties that the consumer realizes a satisfactory article. To put it another way, by passing on this knowledge of the properties of a product, a manufacturer can assure an ultimate consumer that he is getting quality merchandise.

Again to turn from the general to the specific, American Viscose Corporation manufactures rayon and acetate fibers. These products of ours are raw materials to our customer. They are further processed before they become consumer goods.

The relationship between buyer and seller that has been covered is one that exists between American Viscose Corporation and its suppliers. American Viscose feels obligated to carry this relationship on to its customers, such that the ultimate product will have the full benefit.

As mentioned previously, for the customer to get a satisfactory textile product, the correct fiber must be used, design and manufacture of the fabric must be adequate, and the finish or after-treatment of the fabric must be adequate for a specific end use.

We have dealt here primarily with the manufacture of a correct fiber for a specific end use and the approach of passing on the knowledge, gained in manufacture of this fiber, to the yarn and fabric manufacturers and to the fabric finisher.

It has too often been the case that even when the correct fiber has been used an unsatisfactory garment has resulted due to using the fiber

in a sleazy fabric construction or inadequately finishing the fabric. To the consumer the fiber is at fault - not the yarn or fabric or the finish. No better example could be given than the "explosive" sweaters of a few years ago. Early conclusions by the consumer placed the blame on the fiber and not the fabric construction and finish the fabric had received.

Our obligation of passing on complete knowledge of our product encompasses the desire that not only is our customer satisfied with the product and service that he gets, but that the ultimate consumer is satisfied. To do this we work with out customers, that is, the manufacturers between us and the ultimate consumer. We say to these manufacturers that if materials or fabrics are produced from our fibers such that a satisfactory consumer product results that we will assist in the promotion and advertising of these fabrics to the ultimate consumer. Symbol to the ultimate consumer of our approval of the fabric is the Avisco Integrity Tag. To the customer this tag, which is found on the fabric, rug, or garment, means that the fabric used will perform as claimed.

In this program American Viscose manufactures the basic fibers. Since these fibers are man-made, they can be made with a variety of properties, that is, tailor-made for specific end uses. We sell these fibers together with our knowledge as to how they will process on our customer's equipment, and our knowledge and recommendation of textile end product design. All of this is aimed toward providing the customer with a product with which he is satisfied. In doing this we promote the product that carries the label of the Integrity Tag. This means that the customer will have faith in the quality and satisfaction obtained from textile materials bearing this tag. A friend has been made both for the manufacturer of the consumer product and for rayon or acetate made by American Viscose.

Now what does this tag signify? It testifies that properties of the materials used have been evaluated and found to meet minimum requirements for satisfactory performance for the particular consumer product. These requirements obviously vary for the product involved. Fabric serviceability in relation to end use is the basis of the Avisco Integrity Plan requirements. These requirements of serviceability cover the important properties of the consumer product.

Actually, in the textile apparel and home furnishing, the customer decides on an article first on its appearance, its aesthetic appeal. The dress has the desired color and shape, or the rug looks nice and again its color and shape are satisfactory to the eye.

Secondly, the customer usually decides on an article by its feel. That is, it has the proper handle and "pedal". Permit me the use of these words to include feel by both hand and foot. We appraise rugs by stepping on them. These two factors in an appraisal of a textile consumer product are sensory ones.

After the sensory decisions have been made, the consumer of textile materials then is concerned generally with the permanence and restorability of these values. In detail, these questions could be asked: Is the fabric strong enough for reasonable service? Will the strength be adversely affected by treatments it may receive? Will the size or shape of the fabric change - will it shrink? Are the colors permanent? Will

the colors stand exposure to light, rubbing, laundering or cleaning, atmospheric exposure? Will the rug retain its original appearance? Is the rug density enough such that long wear and a good springy feel are assured? Will the tufted bedspread shed lint? Will the blanket mat down or remain fluffy? When the material becomes soiled, how should it be handled to restore the original sensory values?

In general, we are all capable of making the sensory decisions involved in deciding on a textile material. On the other hand, how does the ultimate consumer evaluate the capability of this textile material to retain these values? The consumer can purchase an article strictly on appearance and feel with no knowledge or assurance of its serviceability and, as such, obtain her first knowledge of permanence as the article is used. Sometimes this approach can be quite disastrous as was experienced with the "explosive" sweater already mentioned. When the textile article is purchased at a reputable store, the customer may get satisfaction if she returns the article when dissatisfied with its performance.

One way for the customer to know about the fabric serviceability is for the customer to learn at the time of purchase how the material may be expected to perform and the care that is necessary. The garment or rug manufacturer's name label alone creates confidence in the merchandise but, unless that label is used on a single product with one end use, it is limited in the inference that can be drawn as to the serviceability. When this information is known from service or performance testing and laboratory testing, then the information can be passed on to the customer in the form of a tag or label on the material. Thus by mere inspection of such a tag the customer is assured as to the serviceability and informed as to the care that should be used. Such a program of testing and labeling is the Avisco Integrity Plan. By advertising and promotional work the customer has the assurance that he has found a reliable guide to the purchase of fabrics and textile materials that perform well.

From a variety of different rayon and acetate fibers we recommend to our customers, yarn and fabric manufacturers, fibers of specific properties. We make available our knowledge of the fibers and how they might be handled such that their properties are most beneficial to the manufacturer in making a fabric or textile material. Under our careful coaching fabric dyers and finishers improve their know-how of fabric quality control. Through a license agreement they are bound to comply with the requirements of the plan.

The Integrity Tag is awarded only to fabrics that have met or exceeded test requirements. Among the standards used are the "American Standard Minimum Requirements for Rayon and Acetate Fabrics", known as ASA L22", published by American Standards Association and recommended by the National Retail Dry Goods Association. Use is also made of American Standard L24, "Minimum Requirements for Institutional Textiles". This standard was developed under the sponsorship of the American Hotel Association and is included in their "Hotel Textile Purchasing Guide". Fabrics which we are satisfied will pass the test requirements and which, in general, are manufactured wholly from Avisco fibers are eligible to carry the "Avisco" Integrity Tag or label through every branch of the trade and finally to the customer.

Fabric converters and finishers are required to submit to us samples of fabrics that are to carry the Integrity Tag. These samples are

tested to see that they meet the above minimum requirements for a particular end use.

Basically there is one point of view, whether as manufacturers or as ultimate consumers, we are all purchasers and users. Just as the manufacturer buys wood pulp and uses it to make rayon, the customer buys a rug and uses it to enhance the beauty and comfort of his home. We buy a particular type of wood pulp with the assurance that we can use it satisfactorily. The ultimate consumer wants to buy a rug with the assurance that it will perform satisfactorily. For varied reasons there are limitations as to the customer's ability to appraise the properties that mean satisfaction before the material involved is used. In the same manner that knowledge of a raw material enables a manufacturer to make a product, so knowledge of a consumer product in the form of labeling enables the consumer to purchase with assurance that she will receive satisfaction.

GUIDE TO THE USE OF STATISTICS
IN THE CHEMICAL INDUSTRY *

R. S. Bingham, Jr.
The Carborundum Company

Introduction

Before we can decide whether statistical methods are in order, and if so which ones, it is important to consider the process under study. In chemical manufacturing certain characteristics are important to a quality engineer:

1. There is usually a series of steps involving physical and/or chemical changes; using

2. a variety of raw materials from natural products, to finished chemicals (ore, sand, clay, air, coal to gasoline, coke); in which

3. control requires understanding of

   (a) the properties of materials under reaction
   (b) material handling techniques (barge, pipeline, carboy, etc.)
   (c) unit operation or process characteristics
   (d) measurement method used in analytical labs.

Where to Start

1. Most logical place - control laboratory

   a. highly repetitious operations
   b. need for calibration
   c. standardization
   d. check on sampling
   e. influence of results on other studies

2. Most profitable long range - research and development

   a. the inferences made can be as broad as possible
   b. experimental results can be compared with outcomes of other test programs
   c. results can be measured against estimates of experimental error associated with the test rather than some ideal or guess.
   d. results will suggest next areas to study.

3. Greatest immediate return - process control

   a. leads to distinctions between natural fluctuations and reaction variations accounting for material losses or lower yields, etc.

---

*A more complete paper entitled "Guide to the Use of Statistics in the Chemical Industry" has been accepted by Industrial Quality Control for publication.

b. verifies meaningful "rules of thumb" -
c. reduces number of "adjustments" necessary
d. focuses attention on what the process actually
does and which factors control it.

Power of Statistical Methods

1. statistical "tools" as working implements and means
to an end.

2. general versus special purpose tools

3. design capabilities, limitations in use

4. early disappointments

a. confidence limits too wide; underestimating experi-
mental error
b. not "cure-all"
c. don't replace technology or science skill

5. statistical methods don't create original or final
conditions, rather

(a) measure
(b) point out size of variability
(c) suggest correlation

Questions Indicative of Application

See Tables II - IV

Table I suggests possible use by organizational need.
Table V lists additional sources of information.

Table I - Classification of Statistical Usage By Organizational Function

Statistical Tool	Research Development, Process Improvement	Process Control	Control Lab.	Process and Equipment Design
descriptive statistics	X	X	X	X
frequency distributions	X	X	X	X
control charts		X	X	
significance tests	X		X	X
analysis of variance	X		X	
experimental design	X	X	X	
multiple regression	X		X	X
simple regression	X		X	X
curve fitting	X			
tests of normality	X			
acceptance sampling			X	
analysis of counted data	X	X	X	
"short cut" methods		X	X	
process capability		X		
scatter diagrams	X		X	
confidence limits	X	X	X	
tolerance limits	X	X	X	X
Hotelling's generalized T	X			X
covariance	X			
discriminant functions	X			

161

Table II - General Problems Suited to Statistical Treatment in Research and Development

Problem	Analysis	Statistical Approach
1. How should experiments be planned to maximize information received for a given cost or effort?	1. Define problem. Identify major and minor variables. Select ranges within which to test. Estimate experimental error. Set goal for improvement acceptable. Set failure risks and consequences. Consider interpretations of possible outcomes of experiment.	1. From risks, set degree of improvement wanted, estimated experimental error, determine sample size (n). Select a design to maximize power for a given sample size. Balance and/or randomize to provide internal estimate of experimental error. Compare design with reality; modify. Sacrifice interactions when necessary for broader design.
2. Which experimental designs are best suited to find optimum conditions?	2. Depends on "response system". If any knowledge of this is available, select an experimental design best suited to it.	2. One-at-a-time method. Box-Wilson steepest ascent using fractional replicates, multifactor designs. Factorials.
3. What methods are best suited for comparing different levels of process operation?	3. Is previous estimate of process variability at hand? How large a difference is important?	3. With an estimate, use control charts. Without, use analysis of variance.
4. Plant variables cannot all be changed at once. What can be done?	4. Plan experiment so smallest expected changes are tested first to verify knowledge of process. Then make other changes gradually. Reset variables as best possible to initial conditions.	4. Use multiple regression to study influence of factors. Use control charts during tests to spot unusual behavior. Avoid confounding variables with time wherever possible. Use "Evolutionary Operations" method.

162

5. Several apparently extraneous factors may have influence, but this is uncertain. How do you measure and correct for this?

5. Take measurements on them simultaneously with variables under study.

5. Use covariance methods to determine if they are significant factors. Covariance essentially adjusts data to average of concomitant factors. If found significant factors, include in next test.

Table III - Process Control Uses of Statistical Methods

Statement of Problem	Analysis	Statistical Approach
1. How should a process be controlled?	1. What are the process variables? Which are strongest? Which are measurable and controllable? What points in process are most sensitive, most representative?	1. Make "process capability" study of strongest variables at points in process to verify knowledge and establish state of control. (use individuals control charts, moving range charts)
2. Will a process meet specifications?	2. Are the specs. well defined? Are test methods available for all parts of specs?	2. Is process in statistical control? at right level? with right dispersion? (control charts: normal area)
3. How often should the process be sampled?	3. What are the consequences if it changes from standard values? What does it cost to sample and test? How often is it likely to go out of control?	3. Economic balance of value vs. cost. Power of control chart vs. sample size. Auxiliary use of "runs", "▭ limits", extreme value theory.
4. How should a bulk sample be taken and used?	4. What is known of its history? Is material stratified? Is average or dispersion of interest? How variable is the product? Can the process be sampled prior to bulking?	4. Analysis of variance to establish sampling variability. Compare various ways of taking sample to establish size of stratification effect.

164

Table IV - Control Laboratory Uses of Statistical Methods

Problem	Analysis	Statistical Approach
1. Describe a procedure to verify that routine lab methods are reliable.	1. Reliability requires the state of statistical control and implies that adequate accuracy and precision are present.	1. Use control charts to test for control. Establish accuracy in comparison against a standard. Verify precision from standard deviation.
2. How can lab processes be monitored to maintain control.	2. Comparison against standard performance is implied. Necessary to establish process capability, then compare performance with this.	2. Submit "blind" samples as checks. Plot on control chart as difference from lab analysis. Analysis of variance to establish inherent variability of method.
3. Should a submitted sample be analyzed in replicate? as a composite or in duplicate?	3. What confidence limits are wanted on the mean result? Is a sample of 1 sufficient?	3. Use analysis of variance to identify size of residual errors, sampling, and processing. Note that duplicates are for gross blunder control only.
4. How can product acceptance information be used in process control?	4. Establish correlation of product characteristic and processing variable creating or controlling this. Is product in control?	4. Simple or multiple regression to establish correlation. Control charts. Acceptance plan may be lot by lot variables sampling, or continuous AOQL.
5. Which analytical method is more precise? more accurate?	5. Compare standard deviations for precision; mean with a standard for accuracy.	5. Significance tests; variance omnibus test or analysis of variance.

Table V — Where to Go For Information

Meetings	Short Courses	Journals	Books	Societies
ASQC	Rochester Institute of Technology	"Industrial Quality Control"	Bennett and Franklin "Statistical Analysis in Chemistry and the Chemical Industry".	ASQC (Chemical Division)
- National				American Statistical Association (Section on Physical and Engineering Sciences)
- Regional	University of Oklahoma	"Journal of the American Statistical Association"	Davies, O. L. "Statistical Methods in Research and Production." and "Design and Analysis of Industrial Experiments".	
- Local	Tappi	"Industrial and Engineering Chemistry" (Dr. W.J. Youden's column.)		
- Divisional	Rutgers University (Advanced Course)	"Analytical Chemistry"	Brownlee, K. A. "Industrial Experimentation". 3rd Amer. ed. revised.	
AAAS	Response Surface Methodology (Chemical Division ASQC)	"Chemical Engineering Progress"	Youden, W. J. "Statistical Methods for Chemists".	
Gordon Research Conference		"Applied Statistics"	Bicking, C. A. "Quality Control in the Chemical Process Industries" (in Quality Control Handbook, Juran, J.M., editor.)	
	(see IQC for other listings)	"Biometrics"		

FACTORIAL CHI-SQUARE AS A SEARCH TECHNIQUE

Ralph F. Huth
U. S. Steel Corp.

INTRODUCTION:

Observations recorded as ratios, percentages, or proportions, are commonly referred to as attribute or "all or nothing" type data. Transforming such data into normalized forms often is time-consuming and requires considerable statistical knowledge.

Much of the information collected and recorded during processing operations in a steel plant is of the attribute type. Careful examination and analysis of this information even though it may be an "after the fact" or "hindsight" approach, often leads to interesting and useful conclusions.

The method presented in this paper is aimed specifically at the analysis of process data. Granted, it is better to look to the future in the form of designed experimentation utilizing controlled data collection conditions, but this is not always economically possible, and we cannot afford to ignore information which has already been gathered.

METHOD:

Factorial Chi-Square first became known to the writer in 1956 at the Montreal National Convention. A paper presented by Herbert C. Batson [1] aroused sufficient interest to apply the technique to a current problem with encouraging results. Since then the approach has been used on several other problems, with considerable success.

The basic Chi-Square formula generally appears as follows:

$$\chi^2 = \frac{\Sigma(E - 0)^2}{E} \qquad \text{(formula 1)}$$

Where E is the expected occurrence and O is the observed occurrence.

Let us apply this to the following problem. The lost time accidents for a given year were distributed by shifts as follows:

$$\text{1st shift} = 4$$
$$\text{2nd shift} = 13$$
$$\text{3rd shift} = 13$$

Not knowing hours worked per shift or working condition differences, we assume that the opportunity for accident occurrence should be equal for each shift. Translating this into the terms of our problem our expected occurrence for each shift becomes 10. Solving for Chi-Square with 2 degrees of freedom we get the following:

$$\chi^2 = \frac{(10 - 4)^2}{10} + \frac{(10 - 13)^2}{10} + \frac{(10 - 13)^2}{10} = 3.6 + .9 + .9 = 5.4$$

Our hypothesis is that there is no difference in lost time accident occurrence by shifts. Looking in a table of Chi-Square values for

167

TABLE I
FACTORIAL CHI-SQUARE ANALYSIS

STEEL TYPE	A		B	
PROCESSING METHOD	I	II	I	II
SUCCESSES	31	17	26	26
FAILURES	$\frac{4}{35}$	$\frac{18}{35}$	$\frac{9}{35}$	$\frac{9}{35}$
NUMBER OF TESTS	35	35	35	35

$$\text{CHI-SQUARE} = \left(\frac{140^2}{100 \times 40}\right)\left(\frac{31^2}{35} + \frac{17^2}{35} + \frac{26^2}{35} + \frac{26^2}{35} - \frac{100^2}{140}\right) = 14.27$$
3° OF FREEDOM

STEEL TYPE	A		B							
PROCESSING METHOD	I	II	I	II	$\Sigma+$	$\Sigma-$	D	D^2_N	N^2/SxF	χ^2
SUCCESSES /35	31	17	26	26						
STEEL	+	+	-	-	48	52	-4	.114	4.9	.56
METHOD	+	-	+	-	57	43	14	1.400	4.9	6.86
S x M	+	-	-	+	57	43	14	1.400	4.9	6.86
									TOTAL χ^2 =	14.28

$\chi^2_{.01}$ 1 DEGREE OF FREEDOM = 6.64

168

2 degrees of freedom we find Chi-Square $_{.05}$ = 5.99. We therefore are likely to conclude that there is insufficient evidence for the rejection of our hypothesis even though it may appear that the first shift has a tendency to have fewer accidents.

Supposing our combined results for a two year period had been as follows:

$$
\begin{array}{l}
\text{1st shift} = 8 \\
\text{2nd shift} = 26 \\
\text{3rd shift} = 26
\end{array}
$$

Assuming the same conditions as in the foregoing example we would expect 20 lost time accidents per shift.

The Chi-Square calculation then becomes:

$$\chi^2 = \frac{(20 - 8)^2}{20} + \frac{(20 - 26)^2}{20} + \frac{(20 - 26)^2}{20} = 7.2 + 1.8 + 1.8 = 10.8$$

Our hypothesis that the accidents occur without bias by shifts is rejected at the .01 level of risk. (Chi-Square $_{.01}$ 2 degrees of freedom = 9.21.)

From the two foregoing examples we see that Chi-Square is affected by sample size, just as are most statistical procedures.

Table I gives a simple example of the use of Factorial Chi-Square. The illustrated problem deals with a fatigue bend test wherein a piece of metal is flexed a given number of times and if a crack or breakage occurs, the test piece is regarded as a failure. Two types of steel and two methods of processing are used as the variable factors. The data used is for specimen purposes only.

Using the Brandt-Snedecor [2] formula (Formula 2) the total Chi-Square with 3 degress of freedom is found to be 14.27.

$$\chi^2_{(k-1)} = N^2/(S \times F) \quad X \quad SS \qquad \text{(Formula 2)}$$

where   k = number of subgroups

S = number of successes

F = number of failures

N = total number of samples

$$SS = (s_1^2/n_1 + s_2^2/n_2 + \ldots + s_k^2/n_k - S^2/N$$

Little s and n are the successes and number of samples respectively in each subgroup.

Formula 2 is much faster than formula 1 and yields the same results within rounding error. In this instance formula 1 gives Chi-Square = 14.28 with no rounding error.

TABLE II

2³ FACTORIAL CHI-SQUARE

FACTOR								
A	−	−	−	−	+	+	+	+
B	−	−	+	+	−	−	+	+
C	−	+	−	+	−	+	−	+
S/25	12	13	17	8	8	17	13	12

EFFECT	$\Sigma+$	$\Sigma-$	D^2/N	$N^2/S \times F$	Σ^2
A	50	50	0	4.00	0
B	50	50	0	=	0
C	50	50	0	=	0
AB	58	42	1.28	=	5.12
AC	40	60	2.00	=	8.00
BC	50	50	0	=	0
ABC					

TOTAL X^2 = 13.12

$X^2_{.01}$ 1 DEGREE OF FREEDOM = 6.64

170

The lower portion of table I utilizes the Factorial Chi-Square approach as outlined by Batson in the 1956 transactions. By this method the Chi-Square contribution of each effect is readily known and an independent assessment of the significance of each effect is made.

Table II gives an example of the application of this method to a 2^3 factorial design: an eight cell pattern using the three factors each at two levels. The computations are the same as shown in table I. The interaction which is obvious in table I is not as easily seen in table II until some of the computations are completed. In fact, there might have been a tendency to conclude that there was nothing significant in the data if only a total Chi-Square had been computed, as Chi-Square .01 7 degrees of freedom = 18.48.

This then, is the basic Factorial Chi-Square technique. Next, let us adapt the method to some of the process problems of industry which invariably involve unequal cell sizes when arranged in factorial pattern.

Table III gives an illustration of data of this type - unequal cell sizes. Again the data is for specimen purposes only but is indictative of typical steel problems. The successes represent the band or O.K. edges of approximately 1500 Hot Rolled Coils. The failures are edges that have cracked sufficiently to cause trouble in future processing. The three variables are chemical components of the steel broken into a high (+) and low (-).range. A typical analysis procedure might work as follows:

Step One:

Record all data on Punch Cards.

Step Two:

Find the midpoint of each variable and use this as the breaking point for determining high (+) and low(-) values.

Step Three:

Factorially sort the cards into the eight cells or combinations required by a 2^3 factorial pattern. On a problem with a relatively small number of cells such as in the example, conventional tab sorting equipment would suffice. However, if the number of cells were larger, 16, 32, 64 and on up, an I.B.M. 101 Statistical Machine would do a rapid job and give actual printed listings. 1500 cards could be broken into a 2^5 factorial pattern with 32 cells in less than 5 minutes, provided the board wiring was done ahead of time. However, once a 2^5 board has been prepared additional 2^5 runs could be made with relatively little board wiring time.

Step Four:

Compute the total Chi-Square using formula 2.

Step Five:

Compute the Factorial Chi-Square components using the method

171

TABLE III

2 FACTORIAL CHI-SQUARE WITH UNEQUAL CELL SIZES

A COPPER	-	-	-	-	+	+	+	+
B SULPHUR	-	-	+	+	-	-	+	+
C SILICON	-	+	-	+	-	+	-	+
S	166	118	108	70	110	61	133	88
f	49	32	35	71	70	91	94	202
n	215	150	143	141	180	152	227	290
RATIO •	109	111	106	70	86	57	83	43

EFFECT	$\Sigma'-$	$\Sigma'+$	D'	D'^2	D'^2/N'	$D^2/S' \times F'$	$X^{2'}$
A	396	269	-127	16129	14.30	4.13	59.09
B	363	302	-61	3721	3.30	=	13.63
C	384	281	-103	10609	9.41	=	38.86
AB	346	319	-27	729	.65	=	2.68
AC	315	350	+35	1225	1.09	=	4.50
BC	308	357	+49	2401	2.13	=	8.50
ABC	319	346	+27	729	.65	=	2.68
							130.21

$N' = 1128$ $S' = 665$ $F' = 463$

$N = 1498$ $S = 854$ $F = 644$

$X^2_{.01} = 6.64$

$X^2_{.001} = 10.83$

$$\text{TOTAL } X^2 = \frac{1498^2}{854 \times 644} \left(\frac{166^2}{215} + \frac{118^2}{150} + \frac{108^2}{143} + \frac{70^2}{141} + \frac{110^2}{180} + \frac{61^2}{152} + \frac{133^2}{227} + \frac{88^2}{290} - \frac{854^2}{1498} \right) = 190.86$$

7° OF FREEDOM

• RATIO OF SUCCESSES TO TOTAL OF SMALLEST CELL (141)

172

as shown in table III. In table III the successes are reduced
in the proportion of the cell size (n) to the smallest cell
size, in this case 141. The ratios are rounded to the nearest
whole number. The computations for Factorial Chi-Square are
then made using the ratioed or prime values. N which equaled
1498 becomes 1128, S which was 854 becomes 665 etc. Naturally
our total Factorial Chi-Square is now less than that of our
total Chi-Square because of this manipulation.

Step Six:

It now becomes necessary to interpret and use our results. By
using the Factorial Chi-Square technique we find many of our
effects are significant. However, action should not be taken
just because Chi-Square indicates significance. The Chi-Square
test of significance acts merely as a stop and go sign. If the
sign says stop (no significance) we will not proceed, but just
because it indicates go (significance) doesn't mean we must
take action. In our example the effect of copper was very
strong. Our estimates showed that better than a 20% im-
provement could be made if we worked only the low side of the
copper range. Yet the cost of reducing copper in the steel
might be prohibitive even if it could be done at all. The
sulphur content possibly could be controlled, but would require
some cost expenditure. We must then make a cost study to de-
termine if reducing the sulphur would be economically feasible.
Our best estimate is that an approximate 10% improvement would
result if we worked to the low side of the sulphur range. Per-
haps the silicon effect would be the cheapest to control and
would afford the greatest improvement per dollar spent. The
interaction BC, even though significant might not afford any
increased benefits once the silicon is properly controlled. A
proper examination of a 2-way table should help answer this
question. Because corrective action is difficult or impossi-
ble on some of the effects does not mean we should ignore these
indications in future experimentation, rather, we should use
the knowledge by properly designing future work so that these
effects do not cloud the appraisal of other economically
manageable effects.

USE OF HIGH SPEED COMPUTERS:

Attempts have been made to program portions of the Factorial
Chi-Square technique on high speed computers. Probably the greatest
benefit occurs in the selection of breakpoints for the determination of
high and low groupings. The computers can select the breakpoints so that
the smallest cell size is a maximum. Table IV shows a step by step pro-
cedure for a proposed computer program utilizing Factorial Chi-Square.
The steps of the program are as follows:

Step One:

Prepare a frequency distribution and compute the average and
standard deviation of each variable.

TABLE IV

COMPUTER PROGRAMMING OF FACTORIAL CHI-SQUARE

I   A) FREQUENCY DISTRIBUTION.
    B) $\bar{X}$ OR AVERAGE.
    C) σ OR STANDARD DEVIATION.

II FREQUENCY DISTRIBUTION VS. PERCENT SUCCESSFUL.

III OPTIMUM BREAK POINTS FOR VARIABLES.

IV TOTAL CHI-SQUARE.

V   FACTORIAL CHI-SQUARE.

VI PERCENTAGE EFFECT OF EACH CELL OR POCKET.

174

Step Two:

Plot the percentage of successes for each increment of the
frequency distribution for each variable. This enables a crude
check for serious departures from linearity.

Step Three:

Determine the optimum break points for each variable so that
the smallest sample size will be a maximum. In the actual
running of a problem minimum cell size limits are given the
computer and if this minimum is not met, predetermined varia-
bles are automatically dropped until such time as the original
minimum pocket size is met.

Step Four:

Calculate the total Chi-Square using formula 2.

Step Five:

Calculate the Factorial Chi-Square components using the
smallest cell sample size as a basis for the computations
(same method as table III.)

Step Six:

Calculate the percentage effect of each cell for use in
interpretation of results. The percentage calculations can be
of help in presenting the data to management and can help
answer the question of the percent improvement that can be
expected under various processing condition combinations.

SUMMARY:

Some of the computational details of Factorial Chi-Square are pre-
sented. In addition a technique for handling unequal sample sizes is
presented along with a program outline for high-speed computers. While
the outlined procedures will yield a Chi-Square value that can be tested
for significance, the occurrence of a significant value is not always
cause for corrective action. Tempering the results with good judgement
will enable the experimenter to use this technique on many attribute
type problems. The ease and speed of the calculations should be of
benefit to any experimenter or analyser who is interested in quick ap-
proximations with fair reliability on attribute-type data problems. It
is again emphasized that the Chi-Square test of significance is used
only as a guide, supplementing good judgement.

In using unequal sample sizes the total Factorial Chi-Square sig-
nificance level will always be less than the total Chi-Square value for
the problem. If the total Factorial Chi-Square value is considerably
less than the total Chi-Square value, it is an indication that the sample
sizes are out of proportion to each other. If the sample sizes are ap-
proximately equal and the minimum sample size is 20 or 25, the method
should work quickly and satisfactorily.

175

REFERENCES:

1. Batson, H. C., Applications of Factorial Chi-Square Analysis to Experiments in Chemistry. National Convention Transactions A. S. Q. C. pp 9-23, 1956.

2. Snedecor, G. W., Statistical Methods. Iowa State College President, Ames, 4th Edition, 1946.

WHO CONTROLS QUALITY - AND HOW?

Warren R. Purcell
Raytheon Manufacturing Company

INTRODUCTION

Picture a manufacturing organization with Inspection and Quality Control groups which perform their functions perfectly. The Chief Inspector is a genius at administration, as is the Quality Control Manager. All inspection, measuring, and test equipment is maintained in excellent condition. Inspectors never err, even on visual characteristics. The sampling plans in use maintain just the right balance between cost of inspection and cost of wrong acceptance decisions. The Quality Control group supplies reliable quality data in useful form, and is well staffed with quality control engineers who have unusual ability to diagnose process ills quickly and accurately and make recommendations which, if followed, would eventually eliminate all quality troubles.

Would such a manufacturing organization have quality troubles?
Of course it would.
Why?

The answer is well known and has frequently been stated. It is this: quality is controlled by many persons, and most of them are in neither the Inspection nor Quality Control groups. They are in many other groups, including General Management, Accounting, Research and Development, Design, Planning, Purchasing, Production, Production Control, Industrial Engineering, and Sales.

How can these groups help in the control of quality? What, specifically, can they do?

To help in answering these questions the following suggestions are made. Admittedly, the list is not complete. It is requested that other ideas which occur to readers be submitted in letters to the Editor of INDUSTRIAL QUALITY CONTROL.

GENERAL MANAGEMENT

1. Maintain a consistent quality policy.

   There will be occasional quality failures. Do not let an occasional failure, exaggerated out of proportion because of the particular persons affected, cause an excited drive to tighten up inspection. For this starts the oscillation of the pendulum between quality at the price of quantity and quantity at the price of quality.

2. Keep informed on the total cost of quality.

   These costs fall into three broad categories: prevention, appraisal, and failure. This has been pointed out by Feigenbaum.(1) Count the total, the total counts. An increase in one category associated with a greater decrease in the others is good; a decrease in one category associated with a greater increase in the others is bad.

3. Keep informed on "market quality."

If your quality is poorer than that of your competitors, you are in grave danger. If it is better by a wide margin, your quality costs may be too high.

4. Maintain the balance between know-how and open minds.

Create an atmosphere in which all the know-how in the organization may be used, but do not let opinions stand in the way of finding facts.

ACCOUNTING

1. Maintain adequate records of internal quality costs.

For general management, the breakdown into prevention, appraisal, and failure may be enough. For the quality control manager, however, these should be further broken down by cost centers, parts, defects, and other classifications. For the essence of good quality control management is the concentration of adequate quality control skills at the points of concentration of quality costs.

2. Maintain adequate records of purchased parts quality costs.

The costs of source inspection, detail sorting of lots rejected by sampling, rework by the vendee, and other costs generated by poor quality of purchased parts are not properly chargeable to Inspection or Quality Control; they are incurred by Purchasing. They should be charged to Purchasing, which should, at its discretion, charge them in turn to the vendor.

RESEARCH AND DEVELOPMENT

1. Make full use of modern techniques of experimenting.

Decisions made by Research and Development have important effects on quality and its control. Choices are made among materials, methods of testing them, methods of processing them. Many of these decisions are reached by running experiments. The traditional methods of experimenting are not only costly and inefficient; they also yield wrong answers. The Research and Development group which does not use modern techniques of experimenting is simply putting its company out of competition.

2. Beware of the sensitive process.

A process requiring close control may succeed in the laboratory, but be too sensitive for the factory with its background of more severe variation. The search for a less sensitive process may pay handsomely in reducing the cost of quality.

DESIGN

1. Base design changes on factual feedback of information.

178

Design should be an active part of closed feedback loops, basing its changes on production difficulties, inspection results, and customer reactions. The information from these sources should be based on factual data which have been distilled by a reliable quality control staff.

2. Recognize tolerancing as a major technique of engineering.

While there still remains some degree of guess-work in establishing tolerances, much progress has been made in developing a truly scientific approach to this important subject. The following principles are now generally recognized:

      (a) A tolerance is of no value unless there are means available to verify conformance to it.

      (b) The tolerance capability of a process can be measured scientifically.

      (c) Limitation of designs by the extremes of possible tolerance combinations is unrealistic pessimism.

3. Analyze designs for potential quality troubles.

The Production and Quality Control groups provide valuable feedback information on this problem.

4. Allow an adequate "proving-in" period for new designs.

Do not be misled by a test of too few pieces, or by the fallacies of traditional methods of experimenting.

PLANNING

1. Use manufacturing equipment capable of meeting engineering tolerances.

Do not depend entirely on unsupported claims. Have the equipment accuracy capability analyzed by competent quality control engineers.

2. Use inspection and test equipment compatible with the accuracies required.

Do not depend entirely on unsupported claims. Have the equipment accuracy checked by competent quality control engineers.

3. Before buying new equipment for quality reasons, be sure the present equipment is doing the best it can.

Do not depend entirely on unsupported statements. Have the accuracy capability of the present equipment analyzed by competent quality control engineers.

4. Consider the quality aspects of the locations of inspection and test areas.

> The following questions, among others, arise:

>> (a) Can adequate lighting be provided?

>> (b) Can adequate freedom from noise, vibration, dust, and other troublesome environments be provided?

>> (c) Is receiving inspection near the receiving area? Are inspection stations near their corresponding work flows?

PURCHASING

1. Assume the primary responsibility for the conformance of purchased parts to engineering specifications.

> The costs incurred by lack of conformance can properly be charged to the purchasing function.

2. Keep informed on the costs of non-conformance of purchased parts.

> Get information on these costs from accounting. Use the services of the quality control vendor rating system in selecting vendors.

3. Initiate contracts with qualified vendors only.

> Qualification may be determined in one or more of three ways:

>> (a) Reputation for quality in products similar to the one for which a contract is being negotiated.

>> (b) Experience with the vendor on similar products.

>> (c) A quality control survey of the vendor's manufacturing and inspection facilities.

4. Allow an adequate "proving-in" period for a vendor on a part he has not supplied before.

> In spite of reputation, experience, or an advance survey, the vendor may still encounter difficulties and/or misunderstandings. Advance samples shipped to the vendee for inspection and approval, prior to the start of a substantial run, are advisable.

5. Secure deliveries early enough to permit adequate inspection and a free decision to accept or reject.

> If parts are so badly needed that they must be accepted even if non-conforming, quality trouble is in store.

180

6. Accept, and secure acceptance from vendors, of the principle of lot rejections.

> Lots of acceptable quality should be the normal expectation. The verification of acceptability can usually be done by sampling inspection, at normal quality cost. If the lot quality is not acceptable, it can also be rejected at normal quality cost. The sorting of defective pieces, however, is an excessive quality cost, and should be borne by the vendor, not the vendee.

7. Cooperate in programs of crosschecks of gauges and test equipment.

> Such crosschecks are preventatives of misunderstandings between vendor and vendee. Sometimes the best method of cooperating is by letting the vendor's and vendee's quality control groups set up and operate the programs without interference.

PRODUCTION

1. Provide adequate quality training for operators.

> This is a platitude. But its importance must still be stressed. The difficulty is that immediate problems compete more severely for the supervisor's time than do longer-term problems such as training.

2. Use in-process inspection as a service to production.

> In-process inspection prevents the expenditure of further cost on defective parts. It also provides immediate feedback information for process control. In both respects it is a service, not a deterrent, to production.

3. Do not ask process inspectors to supervise operators.

> It is sometimes suggested that roving inspectors point out directly to operators that their techniques or methods are faulty. This is a serious violation of the prerogatives of the production supervisor. Inspectors should criticize the product or the process, not the person.

4. Do not expect acceptance inspectors to "flinch".

> The function of the acceptance inspector is to check the conformance of the piece to the specification. If it conforms, he accepts; if it does not conform, he rejects. That is the end of the acceptance inspector's function. Reconsideration of rejects should be appealed only to a higher authority. Breaking of this policy is the beginning of loss of control.

5. Use the simpler modern quality control techniques as production tools.

There have been developed several modern quality control techniques which require little statistical skill but are based on pure logic. Operators should use these as tools of their own; not as tools which can be handled only by specialists.

6. Attack the chronic quality problems.

   The sporadic problems compete harder for attention, but the chronic problems are usually the more costly.

7. Call in the specialists as soon as necessary.

   Some of the chronic quality problems do not yield to the shop know-how, even supplemented by the simpler quality control techniques. When that point has been reached, call for the services of the specialists versed in the practical use of advanced quality control techniques.

PRODUCTION CONTROL

1. Consider the quality advantages of long production runs.

   Scheduling quite properly attempts to find the balance between the high inventory costs of long production runs and the high manufacturing cost of short runs. Do not overlook the fact that quality costs, as well as production costs, are lower on longer runs.

2. Consider the quality advantages of large lot sizes.

   Quality costs are lower on large lots.

3. Consider the quality advantages of "first in first out."

   Quality changes occur during production, for both clear and obscure causes. Identification of the corresponding parts is easy under "first in first out", difficult under "last in first out".

4. Allow time for adequate inspection and a free decision to accept or reject.

   If parts are so badly needed that they must be accepted even if non-conforming, quality trouble is in store.

INDUSTRIAL ENGINEERING

1. Include adequate quality clauses in incentive plans.

   Strange as it may seem, there are incentive plans in existence which pay bonus for quantity without regard to quality. Most modern plans, however, do not pay for defective pieces. Nevertheless many of these plans fail to penalize for the inspection cost of finding the defectives. Such costs should also be charged against the incentive; otherwise the company pays for inspection in order to pay the quantity bonus

to the operator.

2. In studying inspection methods, consider inspection accuracy.

The value of inspection should be measured not by the number of pieces going through the inspector's hands, but by the number of defective and non-defective pieces correctly identified.

3. In time-studying inspection, consider inspection accuracy.

The value of inspection should be measured not by the number of pieces going through the inspector's hands, but by the number of defective and non-defective pieces correctly identified. Furthermore, the speed of inspection is substantially affected by the number of borderline pieces on which decisions must be made.

SALES

1. Do not exaggerate quality claims to the customer.

The disappointed customer is the complaining customer.

2. Do not exaggerate quality complaints to the factory.

A comment such as: "Our quality was never as poor as it is now" may stir up people. But without specific, unexaggerated facts, the resulting poorly-directed activity is likely to do more harm than good.

3. Encourage direct contact between factory and field.

It has been shown that direct customer contact by a personable representative of quality control, with the contact controlled by Sales, not only provides feedback of factual quality information of value to the factory, but also improves customer relations.

CONCLUSION

The list could be further extended to include personnel, legal, marketing, advertising, industrial relations, receiving, shipping, maintenance, and other groups. Thus it appears that each person in the organization can make specific contributions to the success of quality control.

Everybody controls quality - and how!

(1) A. V. Feigenbaum: "The Challenge of Total Quality Control." INDUSTRIAL QUALITY CONTROL, Vol. XIII, No. 11, May, 1957, ps. 17-23.

MEASURING VISUAL CHARACTERISTICS

C. A. REED

GENERAL ELECTRIC COMPANY

It is well known that "unless you can measure and express it in numbers ----- you don't know much about what you're doing".

Accurate measurement of thickness, weight, torque, noise, electrical energy, etc., is quite common. There are continuous advancements being made in inspection gaging techniques, many of which are automatic and not only eliminate the inspector but automatically adjust the machine or process. In many instances the prime objective is to eliminate the human element. People just have too much variation and are incapable of being consistent.

The usual visual inspection depends almost wholly on the opinions of people, and still worse, there are usually, at best, only vague specifications, such as "must be suitable for plating" or "must be free of objectionable flaws". Therefore, most of us have been depending on the judgment of people who cannot make consistent or intelligent decisions because they have not been equipped to measure. It is not fair to the inspector, the company, the vendor, or the customer to allow the use of immeasureable opinions. To completely eliminate the human element in visual inspection would be a rather tall order, but there are ways of helping people do a better job. Of course, there is always the possibility of using samples or specimens, but in most cases it requires too many and there is the problem of upkeep due to deterioration and loss.

To overcome this sort of situation we have had considerable success by using a system devised to numerically evaluate visual defects, whether for appearance or for functional characteristics. This system reduces inconsistent inspector's judgment and is simple enough so that people like to use it. It makes for better quality level understanding and communication between vendor, manufacturers and customer. It is readily adjusted to any desired quality level and is economical to use as it is designed around the conventional sampling plans but includes grading by degree of defectiveness.

This system is operated in the following manner:

ACCEPTABLE QUALITY SCORE $\overline{X}_A$ - is established by determining what the customer expects for a minimum score or the percent and degree of defectiveness you are willing to accept.

LOT SCORE $\overline{X}$ - is established on each lot by sampling 20 parts, scoring them by the procedure below on a Scoring Work Sheet (Fig. 1) and taking the average of the 20 individual part scores.

THE SCORING WORK SHEET - is intended to establish a uniform scale of
scoring. Specimens of typical defects in the different score levels
will be useful and should be available for the inspector's use when
practical. It is important that the score given each piece represents
the closest description of the worst defect on the Scoring Work Sheet.

INSPECTION INSTRUCTIONS

 1. Select sample (20) parts at random.

 2. Inspect each part under optimum lighting conditions as
     follows:

     First hold part at arm's length.

     a. Inspect for detail - such as printing on a dial.

     b. Inspect for background defects.

     Second bring part in close.

     a. Inspect for detail.

     b. Inspect for background defects.

NOTE: If any defect is found in the first step it will not be
       necessary to take the second step.

COMPUTE LOT SCORE $\overline{X}$ - as in the following example.

Score	Number of Dials	Score Times Number
5	2	10
4	8	32
3	7	21
2	2	4
1	1	1
-10	0	0
Total	20	68

$$\overline{X} = \frac{Score \times Number}{Number} = \frac{68}{20} = 3.4$$

Lots having a score ($\overline{X}$) greater than $\overline{X}_A$ are acceptable.
Lots having a score ($\overline{X}$) less than $\overline{X}_A$ are rejectable.

The exception to the above is when -10 parts are found. If one
(-10) part is found in the sample of 20 and the score is acceptable
after the computation above, draw an additional sample of 30 parts
and inspect for -10's only. If none are found in the 30 additional,
accept the lot. If any more are found, reject the lot.

If 2 or more -10's are found in the sample of 20, reject the lot
immediately and note the dial score as (0) zero.

186

The example work sheet (Fig. 1) shows a previously established minimum acceptable score of 2.5 for this particular part. The lot on which the work sheet was made, scored 3.4 or .9 points better than minimum acceptable. Therefore, the decision to accept this lot was not made by the inspector's judgment or opinion, but rather by the established value of 2.5. The judgment on the part of the inspector was reduced to only determining whether what she observed on one part at a time fitted the description of, for example, class 3 or class 4 degree of defectiveness. Thus the inspector did not have to decide whether the part was good or bad.

When a situation exists that requires a change in acceptance level, either up or down, a simple matter of adjusting the acceptance number does the job completely and to the exact amount required. It is not necessary to require the inspector to "tighten up a little" or "be a little more lenient".

Establishing the most desirable acceptance score number is, of course, an important part of the system. The same general approach should be used as that used in establishing acceptance in any conventional sampling plan. In this case it not only depends on the percentage of defects you are willing to accept, but the degree of defectiveness of those defects. Often it is just a matter of the proper balance between what the capability of a process is and what is required in the final product.

It is also most helpful to summarize these work sheets (Fig. 1) weekly or monthly and plot graphically the percent of the lots rejected, the average score of the accepted lots and the average score of the rejected lots. This points out not only how bad the rejected lots were but how good the accepted lots were, and indicates trends or changes in the process.

There are many ways in which this scoring by number can be applied, the details of which should be worked out for each particular part or situation.

When setting up one of these plans, keep in mind the general idea of breaking down and describing the visual characteristics and assigning numerical values in a form that reduces inspector's decision making and that results in the kind of protection you want.

VISUAL DEFECT SCORING WORK SHEET

Chrome Plated Die-Cast Bezels	Number Found	Times Score of	Score Times Number
None under close visual inspection.	//	5	10
Not visible at arm's length -- visible only after close inspection.	┼┼┼ ///	4	32
Not visible at arm's length at first glance -- apparent at close inspection -- later visible at arm's length by careful scrutiny -- pits, scratches, light stains, minor flashing marks.	┼┼┼ //	3	21
Same as above except heavier or greater degree.	//	2	4
On first look visible at arm's length with little effort -- blisters, peeling, heavy tool marks, dark stains, dents or heavy nicks.	/	1	/
Any defect clearly visible at a distance of five feet.	0	-10	0

Part No.	
Vendor	
Quantity	
Date	
Insp.	

Total Number → 20

Total Score X number → 68

$$\frac{\text{Score X Number}}{\text{Number}} = \text{Lot Score } \bar{X}$$ → 3.4

Min. Acceptable Score $\bar{X}_A$ for this Part → 2.5

(FIG. 1) Accept ☑ Reject ☐

188

"QUALITY CONTROL IN LOCOMOTIVE MANUFACTURE"

F. H. Howard
Chief Inspector
General Motors Diesel Limited

This presentation will be confined to that part of the plant devoted to the fabrication and assembly of the Diesel locomotive and its structure, and will not involve the manufacture of engines or electrical traction and switching equipment. The rate of production during the period of development of this system varied, generally upwards, from three to seven complete locomotive units per week.

The Diesel-electric locomotive is a self-propelled piece of electric motive-power, carrying its own Diesel-powered DC generating station. Variations in direction, horsepower, speed and pulling effort are achieved in the electrical sense by adjusting excitation of rotating equipment, and by the automatic alteration of polarities and motor hook-ups. These are accomplished principally by low-voltage relays, contactors, interlocks and switches, and furthermore there is a host of other circuits - lighting, starting, charging, and protective among others, some of them AC - which are cut in and out from time to time, many automatically. The locomotive therefore has a relatively complex nervous system of high and low voltage DC, and 3-phase low voltage AC. Its stressed foundation is a plate and structural steel fabrication capable of withstanding 400-ton pull and buff loads. This foundation carries most of the propulsion and auxiliary equipment, as well as a hood and cab to shelter machinery and crew; on certain types of locomotive, the hood and frame share the stresses as a deep box-girder. The swivel trucks carry, besides motors, both brake and spring rigging, with pulling and retarding forces transmitted through their centre bearings to the underframe.

A Diesel-electric locomotive sells for a little over $100 per HP, or close to $200,000 per freight unit, when it includes some of the additional equipment added to our basic specification by the railway. It goes to a customer who carries it on his books as a capital asset; a locomotive fleet is of great intrinsic value to him, he may have secured the necessary funds with some difficulty, and he expects to earn his living by pulling trains with it; that is about the only way he can earn his living, and what's more, since he has bought it to replace a fleet of steam locomotives, many of which were only partially depreciated, he is obliged to look for a handsome return on his heavy investment. He will consequently use the equipment severely and intensely, frequently overloading it, frequently under-maintaining it, and often with relatively untrained people. If this capital equipment becomes unavailable due to our inadequate design or workmanship, or even due to his own inadequacies, he is temporarily prevented from earning his living and develops an antipathy to the failed equipment. All of which is said to emphasize that our company is selling a very expensive article, whose reliability must be the over-riding consideration; in other words, the market has been assessed and found to have certain rigid requirements, with little aversion to a high price tag. The cost of achieving acceptable quality cannot be permitted to interfere when the market will tolerate only perfect performance and will pay for it. I will concede, however, that close attention to manufacturing detail discloses areas where an adequate quality level can be achieved by somewhat relaxed standards of workmanship, areas where the customer is certainly not interested in paying any premium for a

189

quality that he cannot exploit, although our own integrity is frequently demonstrated by maintaining high standards even in those areas.
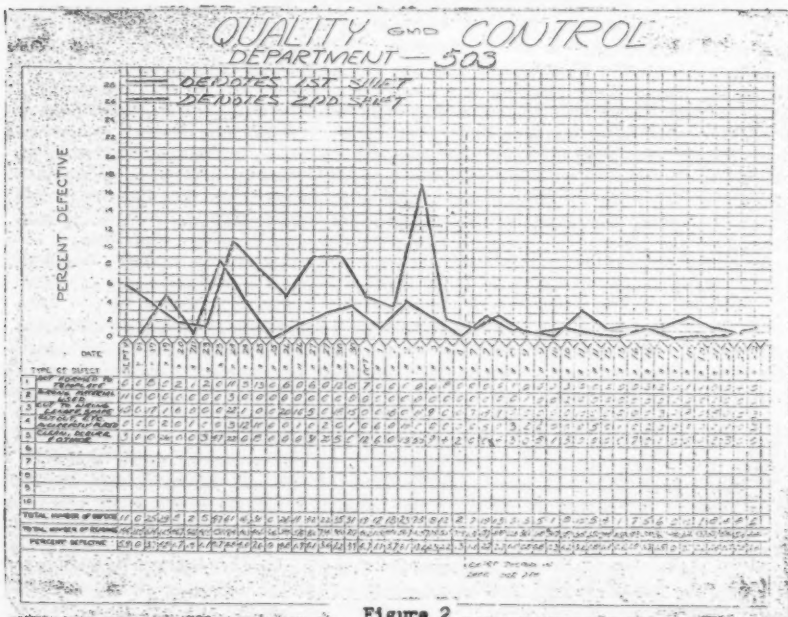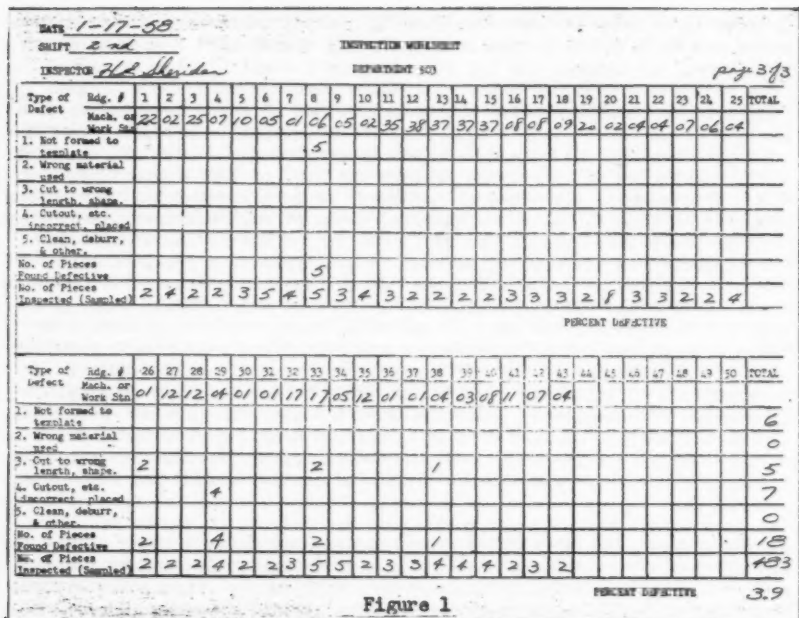
Our locomotives are manufactured on an assembly line, each work station contributing fairly substantial man-hour increments as a unit passes through. Depending upon the number of positions or fixtures in each station, a locomotive may spend from one to five days in one place before moving on to the next. Major components are built on subsidiary lines which feed into the main line at appropriate points. At this time I must describe the system whereby special requests or customer "extras" are introduced into a locomotve. An order for a group of units may include up to 50 extras, accessories that the buying railroad feels are essential to its operation, and for which it pays an additional price. Some of these extras are as complex as the locomotive itself, and are themselves broken down as a matter of efficiency, with portions applied at the various work stations; for example, the piping portions of a dozen extras may be applied in the piping station, along with the basic piping. In short, each station on the assembly line has to know the total net contribution it is to make to a particular order of locomotives, and it may handle units to three or four different specifications on as many successive days. I say net because in many cases extras involve not only additions but substitutions, and therefore deductions of basic equipment.

This group of assembly stations constitutes what are called "consumer departments" because they consume the end-product of the remaining departments, which are known as "producer departments". These latter produce the piece-parts and minor sub-assemblies, in quantity enough for several weeks' consumption, and by part-number, according to a schedule. Their product is not necessarily related to a specific locomotive or order, but does embrace a specific run of various locomotives going through consumer departments.

You will see that quality problems might differ markedly between these two groups. Our original plan of controlling quality was to assign an inspector to each Production Department, on each shift, under an Inspection foreman on each shift. In the producer departments, these inspectors were to sort out good from bad, provide casual in-process inspection when sorting permitted, write scrap or rework tags on the bad, and accept the good; it was our assignment too to count the good, and fill out forms notifying Accounting and Production Control what was to be added to inventory. A carbon copy of this acceptance form was used as a "move" order, (Material Handling would transport nothing without it), and so the inspector had to consult routing sheets to find out and mark down where the various lots of parts were destined to go. Locomotives being compositions of many thousand components, built in orders of from one to 75, and in three or four styles in one mix, these departments produce an extremely high variety of parts, in relatively small lots; so a good proportion of the inspectors' time was spent in the clerical activity, writing perhaps 100 acceptances per shift, and what time was devoted to quality tended to concentrate in the despatch area where nothing much could be prevented except somebody else getting sub-standard material. Prior to the beginning of this narrative, our plant had installed IBM equipment, allowing us to adopt a system of pre-printed acceptance cards, so that all the inspector had left to hand-write was the quantity accepted. This had made more time available for inspecting, and much of it was useful, if rather undirected, "floating" inspection.

Our first move in improving Quality Control in these producer departments was to have the inspector approve the "first-off" a machine or work-bench, or inspect the lot according to a sampling plan extracted from US MIL-STD 105A, or both. Our AQL's were set at various figures judged to be consistent with the contributions of these departments to a quality locomotive: they ranged from 1% on critical sheet-metal to 10% on copper tubing, and were determined by comparing the cost of reworking items undetected by inspection against the cost of 100% inspection. Where pieces moved over several machines, as in a sheet-metal fabricating shop, we arranged to make the sampling check at each machine; finally the deck of pre-printed acceptance cards was put on wheels and accompanied the inspector on his rounds, so when he had approved a group of parts off its last machine he could affix the "move" tag, and go about his business, divorced from the old inspection area, where he had formerly been pinned down for too much of his day. Circulation among the various work stations with pieces being inspected as operations were performed, combined with the abandonment of any sorting philosophy, could only result in a lowered cost of achieving quality, if not a better outgoing quality. Production was committed, under the sampling plan, to sort out good from bad at the machine or work station where the sampling discovered unacceptability, and re-submit their bona-fide best efforts. We devised a homemade kind of "$\bar{p}$" chart which indicated quality trends and kept us all assured that things were under control. Figure 1 shows the form on which was accumulated the data for posting on the charts; it can be seen that we are simply keeping track of the percentage of defects found in the total of pieces constituting all the samples selected in a shift; the machine causing the defect is shown, where applicable. Figure 2 shows the chart for the sheet-metal department, with a line for each shift; the attainment of 4% defective we pronounced satisfactory for a so-called "job shop" operation. Now our Accounting procedure makes provision for scrap to be charged to the department creating it, and rework also to be chargeable to the department causing it, including charges coming back from the department finding the rework and having to do it, wherever it may be; our scrap and rework charges in these producer departments have traditionally been modest, and never cause for serious alarm, so we have never been obsessed by detailed analysis of those accounts. We thought it possibly more than a coincidence however when we did compare charges in three departments before and after our change from sorting after the fact to formal sampling during the process, that they were found in two cases to be lowered by a third and a half respectively. There was decidedly less complaining from consumer departments on misfits and the like; Production was content that better control had been applied.

After several months of satisfactory experience with this new mode, and having duly contemplated the results of a work-sampling study made on these inspectors, we reached for further improvement, and re-organized the activity even more. We have assigned men specifically to all counting and identification, and relieved the departmental inspectors from any responsibility except that of quality; these men do nothing but count parts and handle the IBM paperwork in six producer departments. In some cases this has meant the rebirth of the old "inspection" area in the form of a "despatch" area, since counters cannot distinguish what is acceptable and therefore ready to move, until it has been laid out in front of them. Certain weaknesses may be detected in this approach, but it certainly gets the business of counting away from the business of Quality Controlling, and incidentally lets us see what it costs.

191

DATE: 1-17-58
SHIFT: 2 nd
INSPECTOR: H. L. Sheridan

INSPECTION WORKSHEET
DEPARTMENT 503

Type of Defect	Bldg. # Mach. or Work Sta	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25	TOTAL
		22	02	25	07	10	05	01	06	05	02	35	38	37	37	37	08	08	09	20	02	04	04	07	06	04	
1. Not formed to template									5																		
2. Wrong material used																											
3. Cut to wrong length, shape.																											
4. Cutout, etc. incorrect, placed																											
5. Clean, deburr, & other.																											
No. of Pieces Found Defective										5																		
No. of Pieces Inspected (Sampled)		2	4	2	2	3	5	4	5	3	4	3	2	2	2	2	3	3	3	2	8	3	3	2	2	4	

PERCENT DEFECTIVE

Type of Defect	Bldg. # Mach. or Work Sta	26	27	28	29	30	31	32	33	34	35	36	37	38	39	40	41	42	43	44	45	46	47	48	49	50	TOTAL
		01	12	12	04	01	01	17	17	05	12	01	01	04	03	08	11	07	04								
1. Not formed to template																											6
2. Wrong material used																											0
3. Cut to wrong length, shape.		2							2					1													5
4. Cutout, etc. incorrect, placed					4																						7
5. Clean, deburr, & other.																											0
No. of Pieces Found Defective		2			4				2					1													18
No. of Pieces Inspected (Sampled)		2	2	2	4	2	2	3	5	5	2	3	3	4	4	4	2	3	2								183

PERCENT DEFECTIVE 3.9

Figure 1

QUALITY and CONTROL
DEPARTMENT — 503

--- DENOTES 1ST SHIFT
— DENOTES 2ND SHIFT

Figure 2

The resident inspector has been removed from the pipe shop and sample inspection of individual pipes and pipe assemblies by part number has ceased. Control has been applied to the various processes for attributes which we feel are essential to the reliability of a unit, whose secondary nervous system can be said to consist of piping filled with air at various pressures, hot and cold water, fuel, lubricant, steam and sand. This is handled by a piping inspector whose jurisdiction is shop-wide, and of whom more later; instead of sampling lots of pipes and tubes to find out, for example, which are too long - it will soon enough be discovered if they are too long, or too short - charts have been placed on the flaring, de-burring, threading, soldering, brazing, bending, and washing operations, where proper workmanship could make or break the quality of our piping. These charts are "p" charts with "C" chart information on them; the latter was added because it is essential to collect data as it occurs, since it cannot be duplicated in our kind of shop. The inspector makes his examinations of processes rather than parts, on a random basis, without regard to the particular piece being processed at the time. Being shop-wide, he feeds back to himself and the foreman all the grief encountered on the line, so that corrective steps can be taken immediately.

Resident inspection, sampling all work at all work places, has been removed from both the welded sub-assembly and the sheet-metal departments. From the work sheets associated with our early chart, we had learned which machines in the latter area contributed the most to the defect picture, and control charts have been applied to the ten most significant of them. One inspector now patrols both these departments - the one using up about 40% of the production of the other, and consequently the two side by side - appraising the various operations and operators, and acquainting Production supervision with the contemporary state of affairs. This is carried on with no less confidence than the full-scale sampling, since so little of the product affects the reliability of the locomotive, and what does is well-known. Many of our steel pieces go into fixtures for assembly, where discrepancies in dimension are quickly apparent, and if holes don't line up with other holes, the defect is self-checking. At some small expense of catching defects later in the cycle, we have gained much by reducing costly detailed inspection of low-value parts, that have consistently been in satisfactory control anyway. We hope that we can combine still a third department, the machine shop, into this inspector's area of responsibility, since its kind of operation -   cleaning-up rather than close dimensional work - appears to lend itself to the system.

Classification of characteristics was seized upon in the sheet-metal fabricating shop to lower the cost of achieving quality. This department has been mentioned several times already, and it is in many ways a key one in locomotive manufacture, just as I imagine it would be in aircraft or ship-building. But it seemed to us that it was getting an excessive proportion of our Quality Control efforts relative to the value and importance of its product, so it was not a key department in our operation. Too much time was spent, for one thing, in reviewing and justifying inspectors' decisions, such as whether it mattered if a 2" flange was actually 2 1/16"; a lot of this discussion was by people unenlightened as to actual fact. The drawings of course were quite unequivocal, and called for a routine $\pm 1/32$" on all fractional dimensions (which can hardly be held on general purpose machinery), but like most shops of this nature, work was set-up by and checked against templates. By putting on a

193

"crash" survey program in the consuming departments over a 2-day period, we determined the relative importance of each characteristic on the bulk of basic pieces produced in this shop. The respective templates were then colour-coded to show both Production operators and Quality Control appraisors which characteristics were of critical and which of major importance; they had not known previously. To be precise, we developed a template convention consisting of a series of symbols denoting parallelness, edge distance, squareness, and so on, then painted these signs on the templates with orange for critical ($\neq 1/32$"), nothing at all for major ($\neq 1/16$") and green for minor ($\neq 1/4$"). Figure 3 shows some of these conventional signs. The minor decisions have not formally been exposed to Production workers, although they undoubtedly got to know what the green code meant. This obviates discussion as to whether inspectors' decisions are to be upheld or reversed. Many of us feel that coding of dimensions as to their criticalness, or alternatively setting an appropriate tolerance on each dimension, should originate in the Engineering Department as the design is conceived at the creative source; but it is rather late in the day to start coding drawings in our Company. We think it would be less expensive too, as having it done by a later authority is only refined second-guessing; our Engineers are not unsympathetic to this scheme. At any rate, we have reduced the cost of deciding whether a part is good or bad, as well as the cost of reworking something which was good enough all the time.

Figure 3

So much for the producer groups; let us examine the consumer departments, the major assembly areas. The line consists of work stations where the locomotive underframe is first manufactured in sections; then incorporated into a whole - both principally heavy welding operations; deck-plated; equipped underneath with wiring, fuel tanks, and especially piping; then equipped on top with engine, generator, switchgear, and so on, followed by application of the cab and hood; then set on trucks, put through a stage of final equipping, and despatched to Paint and Test. Major assemblies like cabs and hoods, all equipped internally with their own piping, wiring, and operating devices, are built in their own ancillary areas, as are the trucks. Bear in mind that each of these

assemblies and partially finished locomotives has its own identity, and may differ in much detail from its predecessors and successors. Furthermore, it had become expedient, and still is, for Production to ignore the Bill of Material shown on the drawings, and to rely upon a composite bill supplied by Production Engineering.

In every consumer department, the work it has to do has been laid down, operation by operation, on a check-off sheet; moreover in many cases, each operation has been described separately on its own miniature check sheet, the whole collection of such bits of paper being bound together in a book called immediately and inevitably a "check book". It can be appreciated that here is a vehicle for the introduction of "extras" exactly where they belong, either by the insertion of an additional check, or substituting one for a basic check. Here also is a vehicle for the inspector to make sure everything has been done that his department ought to do, because each check has to be signed off by Production before Inspection will look at that particular increment of work. Furthermore, the entire book of checks has to be counter-signed by Inspection before the locomotive can move on out of the department. Figure 4 shows a typical check; this one calls for the inclusion of a customer "extra", and the drawings describing the entire application are catalogued. All electrical checks have been assigned to a shop-wide Electrical inspector, who demands very high standards on the "primary nervous system". Each inspector, if he disapproves of the work in any way, is to list his points of difference on the back of the check and return it to the Production foreman for a second attempt at doing it properly. Figure 5 shows a typical rejected operation.

GENERAL MOTORS DIESEL LIMITED
MANUFACTURING ROUTING

C.

UNDERFRAME ASM. EQUIPPED

OP. 6916040
18-700
D-443

OPERATION- Apply & hookup the air compressor unloader line.

TOOLING-

INSPECTED BY

COMPLETED BY O2L2.

REFERENCE DRAWINGS- 8160535

Figure 4

At the time this was introduced, it was appreciated that right here was the source material for some rough statistical analysis. All checks were directed to the Inspection headquarters, and defects listed on the backs were there plotted, locomotive by locomotive, and department by department on "C" charts; a simple set of descriptions was appended to break down defects. From time to time a letter would go out to Product-

rectify doors
right side marked

Figure 5

OK 038/007

LOCO ASM. EQUIPPED (before paint)

OP. 6916042
19-540

GENERAL-MOTORS-DIESEL-LIMITED
MANUFACTURING ROUTING

OPERATION- Check fit & alignment of hood doors. Break
tack on hinges and rectify if necessary.
Complete welding all hinges. over

TOOLING-

REFERENCE DRAWINGS- 8208298

INSPECTED BY ___ IN ___

COMPLETED BY ___ OS4 - O45 ___

ion supervision, apprising them of the situations in various areas,
pointing out what kinds of defects had been perpetrated, and from Perado
curves what operation numbers had been trouble spots and what employees
had been principal violators. Later on, control limits were calculated
and instances of break-through reported out.

There were at least two faults with this statistical scheme. One was
that the charts were kept out of sight in the Inspection office area, and
the other was that the analysis covered periods too far in the past. No
noticeable improvement took place as months went by. However, one im-
mediate and quite striking improvement had been achieved. Now that each
operation was spelled out specifically for both Production and Inspection,
the number of minor but irritating defects caught in the final super-in-
spection had dropped by over three-quarters. That was quite tangible,
and justified the amount of work put into preparation and issuance of
the check books, each expended on one locomotive only, in one department.

All checks were filed away in locomotive order and kept for a few
months, so that any defect reported from the field could be traced back
to the check covering the work. Since two people had signed the check
as having seen the work done properly, they could be approached on it.
There turned out to be very little of this necessary however, as our
second immediate and striking improvement was that field-reported or
customer-reported defects almost disappeared.

Figure 6

197

Gradually then, our control charts were moved out to the shop, for display in the various departments, and daily posting. This of course had the effect of bringing them up-to-date and making them of more than historical interest; no more had analysis letters to be written, because the picture was there for all to see, although high-incidence operation numbers and employee identities remained to be exposed. However, that turned out a minor deficiency: just as the "book" said would happen, the day these charts appeared, a trend began - downwards. Men, and foremen, crowded around; rivalry was created; a defensive attitude - not usually desirable, but this time a sign of challenged pride - grew up. Figure 6 shows where one foreman plunged in and started his own statistical analysis, a sort of moving average, and his coloured pencil worked overtime highlighting. It is a commentary on this kind of manufacture when I explain that the sudden upward jump near the end occurred when a major change had been introduced, and the foreman was temporarily distracted from Quality by other problems attendant to it.

Sloping control limits have been applied on an idea borrowed from an aircraft plant. With only a handful of defects in a complex of 200 man-hours, who wants to know the operation numbers or the men accountable? With a final turn of the screw perfection is in sight! Our Inspectors, I daresay quite naturally, tended to become a little more critical in a resolve to prevent utter perfection. Various conflicting ideas of criticalness are a source of trouble, incidentally, in the inspection of a locomotive.

Elsewhere, improvement came in a different way. Figure 7 shows a chart of defects in a department where it seemed difficult to get matters under control. At the suggestion of Production management, who felt that this one chart failed to direct attention to trouble-spots, separate charts were displayed at each of the three or four progressive work stations within the department, so the principal offender could be devoted extra supervision. The small charts' figures were totalled and the result can be seen as satisfactory; the master chart is now broken down by work stations, instead of by attributes.

All these charts were, and are, kept up by the shop inspectors, who early became admirers of the technique, and required only instruction to put us in the fortunate position of not requiring anybody to go around and post charts. The final inspector, the eagle-eye who audits all the other inspectors, was given a pad of forms on which he registered the defects he found and the source whence they came; each preceding department had its control chart modified by what this "mopper-up" discovered and turned in. We had forecast that, based on prior experience with operation and worker numbers, these departments would require an analysis of their defects in more detail than the charts would show; so their own inspectors had these pads too - a sheet is shown on figure 8 - and the kind of defect was coded numerically. This form was used as a work sheet for the chart. Having dispensed with a "chart poster" and analyst, as was just mentioned, there was nobody to sort out the various classifications of defect, and it was accordingly arranged with our IBM group that information on these sheets would be punched onto cards, and tabulated by shift, employee, operation number and nature of defect, whenever we felt we wanted it. But this has never come to pass; with the exception of one department, whose chart you just saw come into control, the situation has never warranted this analysis, and our forecast has, so far, been wrong. We would just as soon it stayed wrong.

Figure 7

Figure 8

Our final refinement has now begun, after several months of acclimatization to this reasonably professional system. Having been assured that quality can be maintained at a high level, we have looked for some profit by reducing the expense of so maintaining it. Our check-book and check sheet system seemed to lend itself to a very careful coding of every operation according to its importance. In six consumer departments, therefore, each operation has been consigned to one of three groups. The most critical ones have been flagged as mandatory check points, where inspectors must witness if possible, but certainly qualify upon completion, the worked called up. An example of an operation requiring mandatory inspection is the alignment of the Diesel engine and the traction generator; 100% inspection is laid down without relaxation. Such checks are surcharged "M". Certain operations where poor workmanship can affect the performance of the locomotive, or can cause the customer trouble or unjustified maintenance, or even lower our reputation in his eyes, are surcharged "Q" and require 100% inspection, but not necessarily at the time the work is done. The remainder are surcharged "S" and cover work that received casual or sampling inspection at the same time that "M" or "Q" inspection is carried out; these jobs invariably are of the kind that have to be done properly before the next job can be started, or can be done only one way. Electrical checks are, as before, to be looked after by the Electrical inspector on a "Q" basis. We have now an Inspection Specification, as it were; little black books and private information are unlamented in their departure. On this philosophy, which contains no measurable element of risk in our opinion, we have been able to reduce the manpower in the consumer departments, retaining the final inspector because the value of our outgoing product still demands him; however, he performs his share of the coded checking. The inspector at the piping position is the general piping inspector, referred to on page 5, and carries out the air test on the piped underframe, a four-hour soap-sud immersion that cannot be dispensed with except at the high cost of finding and repairing air leaks at the Operational Test station.

One of the unwritten assignments of most inspectors, generally assumed by themselves and not seriously challenged by supervision, traditionally has been to carry out simple investigations when things don't go together properly in their areas; such investigation used to lead back to drawing mistakes, use of wrong blueprint revisions, template errors, previous inspection errors, and so on, but seldom took inspectors off the production floor, and prolonged inquiry could not be proceeded with because their presence was necessary for full-scale process inspection. When necessary, Production Engineering used to be brought into the affair. With all-day resident inspection no longer in effect, a pair of "Problem" men has been assigned to take charge of the entire range of this research, and Production Engineering's participation has been greatly reduced. Due to the wide scope of this overall Quality Control, we are contributing towards a real net reduction in Quality Cost. These Quality Control men maintain all departmental charts in addition to being on call by Production or Inspection when fabrication problems arise.

Coincident with our best shop efforts at Quality Control is our feedback of information from the customers' properties. There is really no time limit on the receipt of this intelligence, and 8-year old (million-mile) units developing ailments traceable back to our design or workmanship occasionally receive mention in our Quality Reports. However, the first few months of employment usually show up any defects attributable

to the shop operations as covered in this presentation. These are reported by our Service Engineers engaged in continuous customer contact and inspection of units in service, and are presented to Quality Control for their review. If these deficiencies are in connection with something we are competent to comment on, or to do something about, we enter our remarks on the Quality Report; if one of them has arisen on something made in our shop, interested departments are designated, and where applicable the number of the check book or check sheet operation. The reports are then duplicated and distributed to Purchasing and Receiving Inspection for vendor attention, and to Production and Inspection for shop attention; anything for Engineering scrutiny, or susceptible to design change is relayed to that authority. The departmental Control chart involved gets disfigured by a big red spot at the point where the offending defect has been allowed to escape into the field; as described previously, the check is still on file, and the subject can be discussed with the foreman, inspector, and workman. Figure 9 shows a typical Quality Report with the department and check number shown, as directed to the shop floor. The advent of the check book seemed to coincide with the virtual disappearance of Quality Reports on shop workmanship, and only one or two are coming in each month.

CONFIDENTIAL   DATE July 26, 1957

QUALITY REPORT SUMMARY No. #808

INDEX	REPORT	COMMENTS		
		SERVICE	INSPECTION	ENGINEERING
G57 EB376 CP 8666 GP9	Air Compressor Drive Shaft Nut Loose 803976 2. This nut which helps to secure the drive shaft coupling to the air compressor shaft fell off and was running around inside the coupling. Fortunately this was discovered before the coupling had moved or any damage to the threads had taken place. 7010 Calgary LESS THAN 3500 MILES	PRODUCTION INSPECTION DEFECTIVE WORKMANSHIP!	518 C.B.R. 696040-550	
				QRS No. 808

I.E.D.   CONFIDENTIAL

Figure 9

This is the situation as it exists in General Motors Diesel Limited. We feel that high-priced piece of machinery, assembled to a great extent by hand and with a minimum of automatic machines or processes, requires an Inspection operation that probably can never fall below a certain minimum. We, however, are also convinced that the best Quality Control that we know how or can find out how to apply, will unquestionably result in Production attention to workmanship standards approaching quite closely what detailed Inspection has enforced in the past. It's just as easy to make it right the first time, it's obviously the cheapest way, and training for prevention rather than sorting is, in our kind of business, the professional route to Quality.

AN APPROACH TO QUALITY CONTROL IN AN AUTOMOTIVE PLANT

F. A. Stewart
Chrysler Corporation

This paper deals with what I believe to be the fastest, ficklest, frustratingest but most fascinating business in the world —— the automobile business. I know of no other business where you have a completely new or changed product each year and usually a completely changed plant or facility in which to build them. That is not even to mention the usual running changes and series of crash programs during a model year, which in themselves are enough to keep things off balance. When you combine these conditions with Quality Control, which is a relatively new and rapidly changing field, it can make one wonder about his point of reference from which to work. Having been a Product Engineer for a number of years and a relative new-comer to the field of Quality Control, I can perhaps appreciate these thoughts a little more than many people who are more experienced in this field.

At the Chrysler Division our approach to Quality Control is as practical and down to earth as we can possibly make it. For this reason, rather than attempt to present anything technically new or profound, I have chosen to give you some of our thoughts and experiences in attempting to establish a better organized approach to Quality Control in an automotive plant.

We have been at this business for five or six years at the Chrysler Division, to a more or less degree. We have had many notable triumphs and about as many dismal failures. Looking back over the past and contemplating the future, we think we can catalogue many elements or ingredients in a program or phase of a program which will spell success or failure.

Scope of Operation

At Chrysler the types of operations to be found range from a very highly automated and elaborately tooled machine shop with relatively few direct labor people (Figure 1), to an elaborately tooled sheet metal and
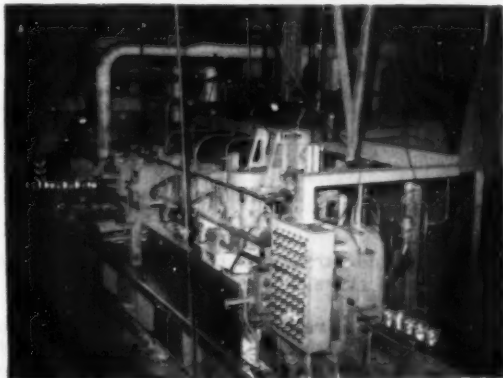

Figure 1

203

body shop with relatively more direct labor people (Figure 2), and finally to the most elementary type of hand assembly operations (Figure 3).

Figure 2

Figure 3

Of course there are a wide variety of operations with all shades of gray of the elements mentioned above.

The materials processed range from castings and forgings in the machining operations through stampings in the sheet metal and welding operations to such things as cloth and soft trim and paint. In addition to the above, tremendous quantities of glass, rubber, adhesives, bright

metal parts and vendor supplied sub-assemblies are processed in the assembly operations.

From this brief description of the operation it probably is quite apparent that the approach to quality control would of necessity take many forms, with techniques and applications tailored to various materials and processes. The fundamental theme of defect prevention remains the same in all areas, however, which invariably involves the determination of the cause and effect relationship with resultant removal or control of the cause.

Organization

The organizational structure which is currently employed or is in the process of being established consists of two essentially parallel organisations. One is Inspection under a General Chief Inspector and the other is a Quality Engineering Group under a Supervisor of Quality Engineering, each of which reports to the Quality Assurance Manager who in turn reports to the Chief Manufacturing Executive. The Inspection section is in the process of being organized along three basic lines; Receiving, In Process, and Final Car Inspection. The Quality Engineering Group is organised similarly along the lines of Receiving and In Process with the third area being Quality Costs.

The functional reporting relationship would then look like the following chart: (Figure 4)
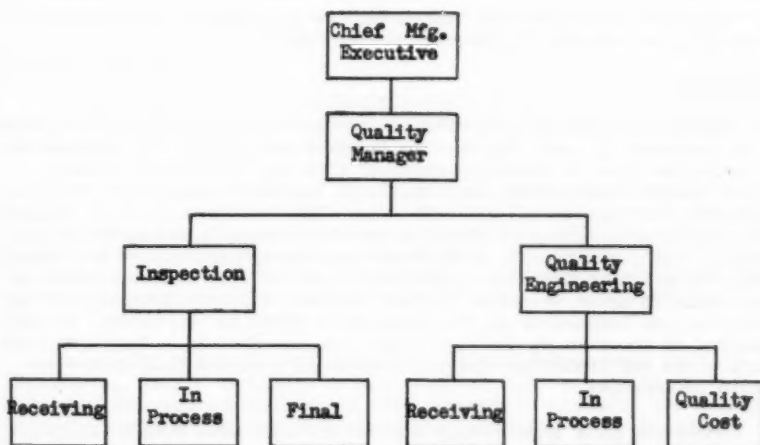

Figure 4

205

Inspection is basically the data gathering and fact finding agency as well as carrying on the routine inspection functions. The Quality Engineering Group is the planning, data analyzing and special study group. There is considerable effort exerted to prevent either of these groups from becoming overly dominant or subservient of the other but instead to have each of their efforts complement and augment the other's.

I might mention at this point that the Quality Engineering Group is an outgrowth or an expansion of a Statistical Quality Control Department we have had for a number of years. By reinforcing this area with more highly skilled or trained people but seasoned with older men with broad experience, we seem to have arrived at the best combination to make the group most effective and respected. One serious hazard I believe has been avoided, that is to have the Statistical Quality Control Department looked upon by the remainder of the organization as a slot in the wall through which you pass a group of data and after waiting a reasonable length of time the data somewhat regrouped and analyzed comes back through the slot for whatever use anyone wishes to put it. I believe we have also reduced the possibility of another facet of the same hazard occurring, and that is to have the Quality Control people become too intrigued with techniques and systems for technique's and system's sake that the sight of the objective which we are all after is entirely lost. We are continually trying to create the feeling and spirit in all people concerned that the only justification of the existence of Quality Control people is, through the use of their training and tools at their disposal, to help production build a product at an acceptable quality level at a competitive cost.

Briefly, the duties and responsibilities of the various areas of Inspection and Quality Engineering are as follows:

Inspection

Receiving Inspection is divided into six major areas which is primarily dictated by plant layout. These areas are pretty well devoted to a particular type of commodity received such as: engine and chassis parts, chassis electrical, body stampings and sheet metal, hard trim and hardware, body electrical and soft trim. Each of these areas is charged with the responsibility of carrying out a pre-determined inspection and sampling plan, maintaining performance records by part number and vendor name. In addition they must coordinate their efforts with the Material Department in order to see to it that material is moved promptly through Receiving and Inspection to the lines or returned to the vendor, sorted, repaired or reworked as the particular case may dictate. They must keep vendors and our Production Control Department informed with up-to-the-minute information.

Receiving is a tremendously big job and must move very quickly. Some idea of the complexity of this job is evident from the facts that at the Chrysler Division alone, approximately 8500 different parts are received and processed. During a normal day approximately 1000 lots or shipments are received. This is roughly equivalent to 350 trucks and 28 freight cars filled with 3 million pounds of material, the dollar value of which is approximately two million dollars. In the interest of controlling the investment in inventory, the average days' supply of material on hand is approximately five days. On many items, particularly the costly ones, the supply is often measured in terms of hours rather than

days.

At the present time the Receiving Inspection areas make use of the MIL STD 105A sampling tables. Eventually we plan to expand this to full usage.

Production Control and the vendor are notified immediately of a rejected shipment. It is usually practical, for the burden of replacing the material, repairing, reworking, or sorting of the defective material is placed on the vendor. Quality information at Receiving is also processed through IBM machines and reported to Quality Engineering where monthly vendor quality performance trends are compiled and reported to Purchasing and other interested agencies for guiding our vendor relationships.

The In Process Inspection is responsible for all inspection throughout the plant. This includes the process inspectors working to a predetermined sampling plan for obtaining information which is posted on charts in each line foreman's area as well as inspection at buy-out points throughout the lines. Here again, the information obtained is reported back to Quality Engineering where it is analyzed and appropriate reports are issued.

The Final Car Inspection is a hundred per cent inspection operation on each car after all of the work and previous inspections are complete. The purpose of this operation is to buy each car for the customer as well as providing a good audit of outgoing quality. Hourly and daily reports and communications are issued to all areas of the plant, including Inspection and Quality Engineering on up-to-the-minute quality problems.

Quality Engineering

As mentioned earlier, the Quality Engineering function is broken down into three functions; namely — Receiving, In Process and Quality Costs, with an engineer in charge of each.

The Receiving Engineering function is oriented between product engineering, purchasing, inspection and production. It is responsible for designing and helping to implement appropriate Receiving Inspection procedures and plans, analyzing data obtained from Inspection and issuing appropriate reports, counciling product engineering, purchasing and vendors on vendor quality problems and designing and conducting any special test which may be required.

The In Process Engineering function is oriented between product engineering, master mechanics or tool engineering, inspection and production. The responsibilities of this area are essentially the same as those for Receiving Engineering except as they would apply to the In Process activities.

The third and most recently added Quality Engineering function is that of Quality Costs. A definite need was felt for having a focal point for the analysis and reporting of the various factors which go to make up the cost of quality such as; warranty, scrap, rework and repairs, inspection costs, etc. This function is oriented between service, product engineering, inspection, production and various financial agencies. We have great hopes for better guidance in our efforts through a better or-

207

ganized approach to quality costs.

In addition to the routine analysis and reporting of quality information and the conducting of special studies, the Quality Engineering Group also conducts weekly quality meetings with each General Superintendent and his supervision. Present at these meetings are all of the service agencies such as product engineering, tool engineering and inspection; in order to bring their efforts to bear on quality problems which are encountered by the General Superintendent involved. These meetings have been very fruitful in the solution of quality problems and have done a lot to sell quality control to production supervision by demonstrating at the local level how it can be of help to them.

Key to Success of Program

In our experience we have had it demonstrated so vividly so many times that the key to success of a quality control program can be summed up in one word ----- attitudes. Proper attitudes are so all important and the one ingredient without which no program can be a success. In the final analysis, isn't this true of about any undertaking?

The attitude of Top Management towards product quality, responsibilities for it and decisions which influence it is all important. I venture to say that there are very few company executives who would not nod their head yes when asked if they are for product quality or product quality improvement at reduced cost through quality control. It is often quite easy to overlook the fact, however, that product quality is affected by practically every decision and plan from the product planning and design stage through all of the remaining stages of plant layout, tooling, procurement, manning, production, inspection and shipment until it arrives in the customer's hands ready for use. It is just as easy to overlook the fact that product quality then becomes an integral part of the responsibility of every area of the organization required to carry out these functions, which is practically everyone.

The attitude of manufacturing or production management cannot be over-emphasized. This is perhaps the most important attitude from the workmanship phase of product quality. It's amazing how the manufacturing management's attitude can permeate the entire production organization down to the lowest man. It follows from this, that a healthy attitude by Production supervision and a willing acceptance of product quality responsibilities will mean a great deal at the grass-roots level --- the operator himself.

Of course we can't overlook the attitudes of the quality people themselves. As I mentioned earlier, we are continually trying to create the attitude in all of the quality people that the main reason for their existence is to aid every phase of manufacturing to carry out their product quality responsibilities. It is not good enough to point out that quality is production's responsibility and then do no more than criticize for this responsibility not being carried out properly. The quality people must be willing to mix right into the production and quality problems, objectively analyse the problems, fairly place responsibilities, ask for corrections, and help provide a vigorous follow-up. We have found this has a chain reaction effect which makes production feel as though someone cares about their quality problems and is willing to help them, in turn creating a proper climate for a healthy quality attitude and thus

reducing your quality problems in both intensity and number. As you know, the production people, of necessity, are a hard headed practical lot directly faced with the daily quality problems as well as many others. If they can see some concrete evidence of help in solving their quality problems, their full cooperation is usually easily obtained.

One of the best examples of this in our organization is a works manager who is of the old school "bull of the woods" type. When it was first attempted to install a quality control system in his area he was not too pleased and as a matter of fact, fought it very strenuously. Finally through the quality people getting into some of his toughest quality problems, conducting machine and process capability studies, giving him good information on scrap and quality performance of his supervision, he became convinced that the quality people were giving him help in highlighting the quality elements which were within his control and making sound recommendations as to how to control them. He was also impressed with the way the quality people went after his quality problems which were beyond his control. As a result, he became one of the dominant figures in the program; the results of which have been exceptionally good, as demonstrated by the following charts of per cent defective vs time, as examples. (Figure 5).

The kind of results exemplified by these curves are brought about by close working team-work of production, Quality Control and other technical service agencies. These curves also demonstrate to me that Quality Control has very little glamour connected with it or spectacular things happening as a result of it. Instead, it is a day to day, constantly plugging type of job continually looking ahead at the long term outlook.

The other day, the works manager mentioned above, stopped in my office to discuss a few problems. At this time he commented that if he were to set up a new operation, he would install exactly the same system. Coming from him, this was one of the highest compliments which could be paid the program and exemplified the spirit or attitude which has contributed to its success.

INTAKE ROCKER ARM

CRANKSHAFT
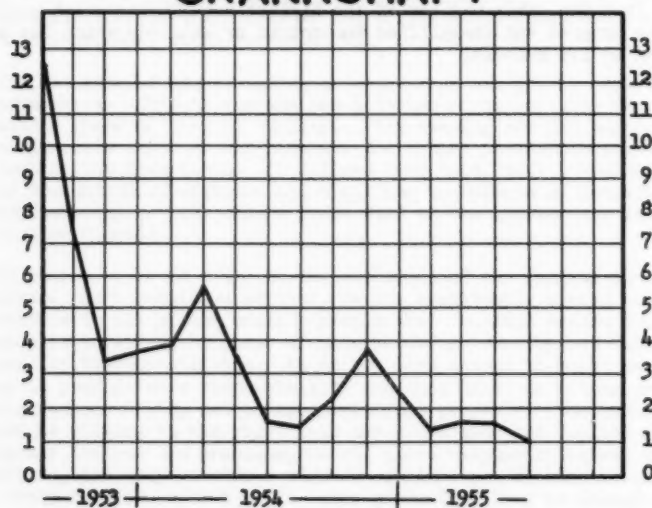
THE MANUFACTURING PROGRESS FUNCTION

Richard Conway and Andrew Schultz, Jr.
Department of Industrial and Engineering Administration
Cornell University

I. INTRODUCTION AND BACKGROUND

    A. Objective of Paper
    B. Manufacturing Progress Defined
    C. Measurements of Progress
    D. Previous Use in Aircraft Industry
        1. General shape
        2. Mathematical description - various bases
        3. Logical basis
        4. Inconsistencies in its use

II. CAUSE AND EFFECT - LOGICAL RELATIONSHIPS

    A. Pre-production Factors
        1. Tooling, Product Design, Process Design, Methods,
           Organization, etc.

    B. During Production
        1. Operator, Methods, Design and Operation Changes,
           Engineering Effort, Volume Changes

III. EMPIRICAL STUDIES

    A. Examples
    B. Inferences
        1. Fits
        2. Prediction

    C. Areas of Application
        1. Firms
        2. Products

    D. Problems of Application
        1. Parameters
        2. Identification of factors - statistical relations
        3. Aggregation
        4. Classification
        5. Points of estimation

IV. CONCLUSIONS

    A. Applicability
    B. Reliability
    C. Further Study

APPLICATION OF THE IBM 650 MDDPM
TO THE MANUFACTURING PROGRESS FUNCTION

William J. Halladay, Jr.
Project Engineer
International Business Machines

I. A VERY BRIEF DESCRIPTION OF THE 650 MAGNETIC DRUM DATA PROCESSING
MACHINE.

The 650 is defined as a digital electronic computer, the
basic units are described and the two primary types of program-
ming are identified.

II. THE USE OF THE 650 MDDPM TO CALCULATE TABLES FOR THE APPLICATION
OF THE MANUFACTURING PROGRESS FUNCTION.

The 650 is programmed to calculate two tables: (1) the X^b
tables where values of X from 2 to 32,000 are raised to fourteen
different b power values, and (2) the E table where a constant E
is calculated for the same values of X and b as in the X^b table.
The X^b values are used in the basic formula $\overline{Y} = a/X^b$ and the E
table is used to calculate M.P.F. hours for given lot sizes.

III. THE USE OF THE 650 MDDPM TO CALCULATE AND EVALUATE THE SLOPE OF
THE MANUFACTURING PROGRESS FUNCTION.

Three integrated 650 programs are used to calculate and eval-
uate the slope of the M.P.F. Program #1 is a Basic Machine Lan-
guage program designed to compile costs per lot by part number
into cost per cumulative unit by unit or machine number. Program
#2 is a "FLOPS" program to consolidate all cost coordinates (from
Program #2) into cost coordinates with significant cost differ-
ences between units of cumulative production. Program #3 is an-
other "FLOPS" program to analyze the results of Program #2. Anal-
ysis is in terms of a Least Squares Trend Line, upper and lower
control limits and the regression coefficient.

214

THE MANUFACTURING PROGRESS FUNCTION,
ITS APPLICATION TO OPERATIONS AT IBM, ENDICOTT

Donald A. Schreiner
Technical Engineer
International Business Machines

I. APPLICATION OF THE MANUFACTURING PROGRESS FUNCTION

   The following application of the Manufacturing Progress
Function will be discussed:

   A. Cost of introducing a new machine or a change to a
       machine.

   B. Prediction of manpower and space requirements.

   C. Budgeting of special engineering and staff efforts.

II. THE TECHNIQUES USED IN MAKING A MANUFACTURING PROGRESS FUNCTION
    STUDY.

   The techniques used for preparing estimates, selecting
"ultimate" units of production, and "progress slopes" will be
discussed in detail.

III. THE RESULTS OF A MANUFACTURING PROGRESS FUNCTION STUDY.

   A discussion of how management is aided by use of the Manu-
facturing Progress Function with regard to personnel transfers,
cost analysis, and planning.

AN EXAMPLE IN STATISTICAL PLANNING OF LABORATORY EXPERIMENTS

Or

THE ENGINEER AND STATISTICIAN CAN MEET

A Skit
in Three Acts

By

The Engineer   F. R. Del Priore
The Statistician   B. B. Day

Engineer's Office

Stat: Good morning, Mr. Del Priore.

Engr: Good morning, Miss Day. Whom did you come to bother this time?
How about a cup of coffee?

Stat: No, thank you, I'v had my cup. You happen to be the victim.
I've been thinking about that problem you were discussing at lunch
last Wednesday. Perhaps I can help you. Would you be willing to go
into the details more completely this morning? Have you the time?

Engr: Oh, I always have time to talk about my baby, but I don't see
how you can help. I know you have helped Mike and Larry, but my
problem is different.

Stat: Where have I heard that before?

Engr: And, besides, the test has not been run yet. Wait until I get
the data, then I will be glad to have some help. In fact, I will need
help because we are going to have reams of it.

Stat: Believe it or not, it's before you get the data that we can be
of most help; I mean in planning the test schedule.

Engr: Oh. Besides, we don't have any money in the job order for your
services.

Stat: That's something you don't have to worry about, fortunately.
The Bureau of Ships provides a special fund to run our office, so we
don't have to charge our time to the test set-up. Our services are
free!!

Engr: That is fine. Well then, perhaps you would like to see the
test set-up.

Stat: I certainly do want to, but I would like to get some more back-
ground information first. It would then mean more to me. What,
definitely, is the purpose of this test?

Engr: The long-range program - overall problem - the Navy must reduce
the noise generated by our submarines. With the increase in efficiency
of sonar, this is an important problem. Our phase is to reduce the
noise from the power transmission gears, particularly the reduction
gears. This will include testing different gear geometry, materials,
finishes, etc. That is the long-range problem.

Stat: Is this something new? This kind of testing?

Engr: Many aspects of it are new. The immediate problem is to set up
test conditions for later testing different gears, etc., in the long-
range problem. This is the problem then.

Stat: I see.

Engr: And the real reason is that we simply have to cut down on all the possible test conditions that we could run.

Stat: Just what are the factors that you are considering?

Engr: May I use my blackboard? I always talk better when I use the blackboard.

Stat: Oh, please do.

Engr: The most important factor is direction of introducing the lubricant.

Blackboard
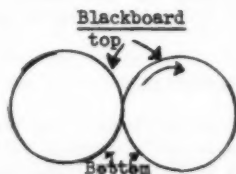
Engr: Here is the gear system and we can introduce the lubricant from the top or from the bottom. We can introduce it in mesh with the system or tangent.

Stat: I see, there are then four possible directions.

Engr: And we have designed a new nozzle to do this. Another factor is speed. With this set-up here we are thinking of testing two speeds - 300 and 1200 rpm. Another factor is load, and we are thinking of testing 25 per cent of full load and 125 per cent. The fourth factor is volume - this is the amount of the lubricant, and we are thinking of 1/2 gallon per minute, 1, 1-1/2, and 2 gallons per minute.

Stat: Are these all - direction, speed, load, and volume?

Blackboard

Direction:	top inmesh, top tangent, bottom inmesh, bottom tangent
Speed:	300 and 1200 rpm
Load:	25%, 125%
Volume:	1/2, 1, 1-1/2, 2 gpm
Temperature:	90°, 120°, 160°F

Engr: Oh, wait a minute! There is one more - temperature, and we are thinking of 90°, 120°, 160° Fahrenheit. Now, these are the factors that we think are important.

Stat: Which is the most important? Direction?

Engr: Yes.

Stat: I see.

Engr: This is going to be run with the standard gears?

219

Stat: What about the lubricant?

**Engr:** 2190-T, standard.

Stat: Oh, are you going to keep this constant for the test?

Engr: Well, that is what the Navy is using all the time - most of the time.

Stat: That sounds good. There are certain things now that we must know to be able to plan efficiently. I'm going to ask a really difficult question: Do any of these factors act jointly; for instance, speed and load? Suppose we plotted our load along here, and we ran a curve for each speed. Let's see, you're going to measure this in decibels, aren't you?

Engr: Decibels, yeah. You straightened us out on that.

Stat: Good. Suppose you plotted decibels along the vertical axis and you ran it at 300 rpm. You would get a certain noise level at the low load and you get another level at the high load. At the 1200 speed at the low load, it may be high and at the high load it may be low. Just the opposite effect. The noise was less. We would say then in the statistical language, there is an interaction effect between speed and load on noise.

Blackboard
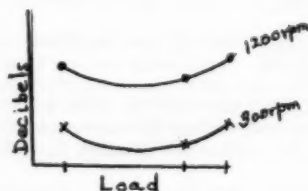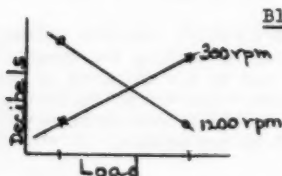
Engr: I don't know what you mean.

Stat: Well, let's try again. Suppose we had a certain curve for this 300 speed and then, for the 1200 speed, it was only higher, but with the same slope. The two lines are pretty much parallel. Then we would say there is no interaction of speed and load on noise. Dr. Youden has explained it in a rather neat way. Your ten-year old boy Jimmy makes a certain amount of noise when he plays in his backyard. Your neighbor has an eight-year old boy who also makes a certain amount of noise in his backyard. Put Jimmy and the neighbor's boy together, and what will you get?

Engr: A racket!

Stat: Then, he carried it a little farther. Suppose you had a 15-year old daughter and your neighbor had a 16-year old son. Each one in his own living room in the evening would make a certain amount of noise, but put them together in the same room......

Engr: It would be awfully quiet!

Stat: That is what we mean by interaction effect.

220

Engr: Suppose you come back to the second graph.

Stat: All right.

Engr: If one speed curve is, say, a little bit higher than the other speed curve but the same slope, then you say it's no interaction?

Stat: That's right.

Engr: Now, it must then be a straight line?

Stat: Oh, no, it might take any shape. Suppose we had three loads. The noise effect might show a curved line; but, if the curves are pretty much the same shape, then we would say there is no interaction.

Engr: I see, it is the same family of curves. Now this means that there is no interaction.

Stat: That's right.

Engr: And you call the first plot: speed and load interact?

Stat: The interaction might take different forms. For instance, if at one speed noise increased faster with an increase in load, we would have an interaction effect.

Engr: Okay.

Stat: Now, let's examine these factors in the light of what we've just said. How about direction with these other factors. Does it interact with any of them?

Engr: Well, so far as direction is concerned, it is new. It is a new factor and we don't know anything about it.

Stat: I see. How about speed?

Engr: Speed is tricky, Miss Day. We have studied speed before and we have found that it, as you would say, interacts. I guess I am using the word correctly. It interacts with these other factors more or less, sometimes; so I don't know.

Stat: Not so good. How about the other three.

Engr: In our previous studies, we have found that load and volume and temperature too are more or less additive to each other. Does that mean they do not interact?

Stat: That's right.

Engr: Additive. If they are additive, that means there is no interaction.

Stat: That is right.

Engr: Then we can say load and volume and temperature do not interact.

221

Stat: Now, let's see what we have here. You have four volumes. Is that so terrifically important that you need four levels?

Engr: No, we just wanted to cover the range from 1/2 to 2.

Stat: Would three be sufficient?

Engr: All right.

Stat: Well, if you left out one, which would you leave out?

Engr: Oh, I think the 1-1/2.

Stat: Good! Now that I have given you something there, how about you giving me another load. Wouldn't it be a good idea to have three loads?

Engr: I guess it would be. How about full load, 100%?

Stat: Good! We have four directions, two speeds - that will make 8 combinations - three loads - that will be 24 - three volumes - that'll be 72 - and three temperatures - that'll make 216 combinations altogether.

Blackboard

$$4 \times 2 \times 3 \times 3 \times 3 = 216$$

Engr: What! 216 combinations?

Stat: That's what you have there.

Engr: And that is without running it three or four or five times? We can't do that, Miss Day.

Stat: Don't get excited; give us a chance. Maybe we can do better than that.

Engr: I hope so.

Stat: We have covered the purpose, the factors, and their levels and interaction. There is one more matter, I think of now, to discuss. How easy is it to change direction?

Engr: Oh, direction is easy, just a flip of the switch.

Stat: I see, and speed?

Engr: Same thing.

Stat: How about load?

Engr: Load takes about an hour to change. Volume takes five minutes.

Stat: And temperature? You haven't said anything about that.

Engr: Temperature. We can't change temperature within a day. We have

to heat up the whole tank of oil or cool it down and we can't do it within a short space of time. I would say it takes about four or five hours to change temperature.

Stat: Well, it is good to know about this. Suppose you had a certain temperature and load and volume. How many could you do in a day of the speed and direction combination?

Engr: Oh, well, we calibrate every morning. Let's see, we can perhaps get eight or ten different runs changing speed and direction if we keep the others fixed - constant.

Stat: That's not too bad. I believe we can work something out.

Engr: Oh, and there is one thing more I forgot to mention. One of the reasons for running different tests is to decide where to take the noise measurements. In this test, we are going to measure the noise at five different locations.

Stat: Only five!

Engr: And then we are going to take tooth contact frequency and also the overall noise.

Stat: Can these all be taken at one time?

Engr: When we have the gear system running, we take all of these at one time. The pick-ups are there. It is on tapes. We don't have to worry about that.

Stat: Well, that won't be any problem then. It just means more data to analyze.

Engr: But that is one of the reasons for running the tests. We hope to be able to cut down on the number of readings in the future.

Stat: I believe I have the picture now.

Engr: You say we have 216 combinations. We just can't run them over and over and over again, Miss Day.

Stat: Well, suppose we don't worry about that until I have time to think this over a little and see what we can come up with. Shall we take a look at your equipment?

Engr: I will be glad to show it to you. You know, this time has not been wasted. Just talking about the problem has helped. If you need any more information, give me a ring.

Stat: Thank you, I may come back for more. It certainly is an interesting problem.

Engr: I am glad you think so. Let's go to the laboratory.

Stat: O. K.

<center>END OF ACT I</center>

Engineer's office - several days later.

Stat: How does 144 or even as low as 72 runs sound for your problem?

Engr: Oh, 72 instead of 216. That sounds real interesting. Now you're talking up my alley. Proceed!

Stat: I'll put it on the board in the form of a little diagram.

Blackboard

Load	Volume 1/2	1	2
25%	120°	90°	160°
100%	90°	160°	120°
125%	160°	120°	90°

Stat: Suppose we put volume across the top. We had three volumes - 1/2, 1, and 2; and load over here on the left side. We had three loads - 25%, 100%, and 125%. Now let's put our temperature in these cells and we'll put one each in each row and in each column. Starting with 90° in the first row and second column; then we put it in the second row and the first column, and finally in the third row and third column. Now, let's see what this means. In the first case we'll run at 25% load, 1 gpm volume, and 90° temperature.

Engr: Is that what it means?

Stat: Yes. Also we'd run the 125% load, 2 gpm, and 90°. Now let's follow through in the same manner for the other two temperatures which will completely fill our nine cells. Thus we now have nine different combinations. If we have made all the possible combinations of load, volume and temperature, we would have had 27. But here we have reduced it to nine.

Engr: Yeah! We have to run the other 18.

Stat: No, not for the present. Let's just think about these nine. Now, let's see what we'll get out of it. Suppose we took the measurements for these nine different combinations and they were put in the nine cells.

Engr: You mean, after the test has been run?

Stat: Yes. You will note that, if you added across and got the total for the first row and the total for the second row, they would be balanced for volume and they would be also balanced for temperature.

224

Engr: Wait a minute, I will have to look at that. (Pause) You are right.

Stat: So that the difference in these totals, aside from the error in running the tests, will measure the difference between using a 25% load and a 100% load. Likewise, if you took the third load and compared it with the other two because they would be balanced throughout. Now we'll write that down. Load - we have two comparisons there, that is two independent comparisons. We can compare our 25 against the 100 and we can compare the two against the third. Now let's look at the column totals which stand for the three volumes. Suppose we took the first volume, the 1/2 gpm, and last column, the 2 gpm. You will note they are balanced for both temperature load. So the difference in the totals over and above the errors in the test will be due to the difference in the two volumes. Do you see that?

Engr: I see. It is balanced.

Stat: Yes, well, then if you take the high against the low, then the other would be the two ends against the middle. So you would have two comparisons for volume. The next step is a little more difficult to see. But look at our temperatures.

Engr: You mean to say that temperature is balanced too?

Stat: Well, didn't we put it in a balanced way in the beginning.

Engr: Well, let's see. (Working on board) 90° here load 1; 90° here load volume 1/2; 90° here volume 2 load 125, 120, 120, 120, 160, 160, 160. It is balanced as of the temperature.

Stat: So, we'll put in two comparisons for temperature. We have nine different runs and there will be eight comparisons. That is one thing you have to take on faith; adding these up and subtracting from the eight, leaves two for the remainder, that is for measurement of the error in the test.

Engr: Error? As I see it here, you only have nine combinations out of the twenty-seven. We are not running any of these over again. It seems to me we are getting something for nothing.

Stat: Well, not exactly. You remember last Monday when we discussed this you assured me there were no interaction effects of volume, load, and temperature. It was for that reason we could use this plan. We call this the "Latin Square". The "no interaction" statement made that possible.

Engr: Latin Square? Wait until I tell my Senator about this. Well, are we going to run the other parts of that twenty-seven?

Stat: Not unless we find we need to. We'll run these first and see how well the data come out.

Engr: Oh! But wait a minute, you have forgotten the most important problem in the test; speed and direction. I mean, that is the reason

225

for running the test.

Stat: I haven't forgotten them. I'm just coming to that part. You had four directions and two speeds, or a total of eight combinations with speed and direction.

Engr: Let's see, we have 300 and 1200 top inmesh, 300 and 1200 top tangent, 300 and 1200 bottom inmesh, 300 and 1200 bottom tangent - 8.

Stat: You're doing fine. Now you also said that you could run 8 or 10 in a day. As you will see in the drawing, what I have done has been to arrange to run 8 combinations of speed and direction for a particular volume, load and temperature in one day. Can they do that?

Engr: You mean we set the volume, load and temperature and then change the speed and direction? yes, we can run eight.

Stat: Fine, so this will represent altogether 9 days of work.

Engr: Right - nine days. Oh, and in this way you take care of temperature. You remember we couldn't change temperature in a day? I was wondering what you were going to do about temperature.

Stat: That's right. Well, now, let's see what comparisons we get with this in the same way as before. We will call the part above the main part, the comparisons for speed and direction we call the subpart. This is a case where the sub-part is more important than the main part.

Engr: It seems to me like a backward way to do it this way, Miss Day.

Stat: Yes, it does.

Engr: Is this the way all statisticians work?

Stat: I guess so. If we have 2 speeds, we can only compare one speed against the other, so we have one comparison for speed, and a very good comparison too for we will have four of each speed in each one of the nine days. Direction - we'll have three comparisons there. We have top against bottom; inmesh against tangential; and then the inter- action of the two. What we also have is the interaction effect of speed and direction. Remember, you said you wanted to find out this. The neat part about this plan is that we can find the interaction of speed with each of the three above also - load, volume and temperature. Also, the interaction of direction with each of them. We will have a total of 72 readings.

Engr: Nine days, eight a day -- seventy-two, right.

Stat: So we'll have 71 comparisons, and we'll have a remainder here, or error term, of 32. You may be interested to know that this sub-part is called "split plot".

Engr: I thought it was Latin Square?

Stat: Well, the whole thing is a "Latin Square Split Plot" design.

Engr: Latin Square split-plot - wait until I tell Mike about this.

226

Stat: Now, the valuable thing about this is that we'll have a very good measure of the effect of speed and direction, also a good measure of the within-day variability or error. Is that all clear?

Engr: Well, frankly, there is a lot about this I don't understand, but I'll try anything once. Suppose I copy down these combinations and we will start running.

Stat: Wait a minute. There's one thing more that I saved until the last. And that is we want to make sure that no systematic changes in your machinery gets into the test. So, what I had done is to randomize the order in which these nine days will be run.

Engr: Randomize?

Stat: Yes, I've actually left it to chance which one would come first, second, etc. In the drawing here, the first one to be run is the one with the 1 gpm, the 25% load, and the 90°. Do you see any objections to this?

Engr: No. We have to set it at something every morning. We can set it at anything you prefer.

Stat: Fine.

Engr: Except, I was going to test all the 90°'s first, then the 120°'s and then the 160°'s. It is the way we always do.

Stat: This is better for it is insurance against some of the systematic effects getting into the readings.

Engr: Okay, and we will set them at whatever you prefer. We can run all the 300 rpm's in the morning and all the 1200's in the afternoon.

Stat: Well, now, wouldn't it be a good idea to be just as careful with the within-days runs to avoid systematic changes we call bias there?

Engr: Well, we know our equipment changes during the day. We calibrate every morning and we know it does not stay on calibration all day long. But I don't see how you are going to take care of that.

Stat: Well, we will do it in the same way by leaving to chance the order in which they are run. And I believe you told me in the beginning that it was very easy to change from one speed and direction to another.

Engr: Yes, but what is wrong with doing it the way we have always been doing it? We will run the 300's in the morning and 1200's in the afternoon. What's wrong with that? It's a nice break.

Stat: Yes, but if there's a change in drift, that drift might coincide just with your changes in speed. I have prepared a random order of runs for the first day. Would you take a look at it and see if it can be done without too much upsetting of the works? (Handing first day's schedule to Engineer.)

227

Engr: I see you mixed up the speeds and directions. You randomized?

Stat: That's right. You're a very apt pupil. We'll furnish a different order for each test run. They're in the process of being typed up now. We have set it up in a way I believe the technicians can follow. Will you see if it needs any revision?

Engr: This part up above is what we set and then during the day we change the directions and the speed and we will follow this order.

Stat: That's right. And then there'll be another one for the next test run.

Engr: A different order?

Stat: Yes. And this is what we call randomization, leaving it to chance.

Engr: This is new to me, Miss Day. This is not the way we usually do it. Let's spend a minute on it. I would like to be clear as to why. If there were a drift and we always ran the 300 speeds first, the 300 speeds would be a little lower than they would be because of the drift?

Stat: That's right. Or a little higher.

Engr: Or a little higher. And the same with the 1200's?

Stat: That's right.

Engr: Oh, I see, so you mixed them up like this, so in the end the picture will be as it should be. Okay.

Stat: Yes. Both speeds have an equal chance of being affected by the drift?

Engr: I hope that you are not going to tell us we should run this over five or six or ten times.

Stat: Oh, no, let's run it this once and analyze the data and see what we get from it. It may be necessary to make a second run, but again it may not.

Engr: This is quite different than what I had in mind, frankly. We will see how it comes out.

Stat: Fine, thank you. That is all anyone could ask - to give it a trial.

Blackboard

Main part	Comparisons
Load	2
Volume	2
Temperature	2
Error	2

Subpart	
Speed	1
Direction	3
S x D	3
S x L	2
S x U	2
S x T	2
D x L	6
D x U	6
D x T	6
Error	32
	71

229

Figure 2 - 3 x 3 Latin Square - Split Plot Design
For Preliminary Gear Test

LOAD	VOLUME		
	1/2 GPM	I GPM	2 GPM
25 %	120°	90°	160°
100%	90°	160°	120°
125%	160°	120°	90°

230

TEST SCHEDULE

First day of 9 days

Preliminary Gear Test for Optimum Operating Conditions

First Day

Main Setting

Load	25%
Volume	1.0 gal/min.
Temperature	90°

Order of testing	Direction	Speed (rpm)
(1)	Tangential - bottom	300
(2)	Inmesh - top	1200
(3)	Inmesh - bottom	300
(4)	Inmesh - top	300
(5)	Tangential - top	1200
(6)	Tangential - top	300
(7)	Tangential - bottom	1200
(8)	Inmesh - bottom	1200

231

Statistician's office (three weeks later)

__Engr:__ Hello, Miss Day.

__Stat:__ Hello, come in.

__Engr:__ It's been a couple of days since you had the data and I wonder if you have it analyzed.

__Stat:__ Well, now, you do expect fast work. We __are__ good, but not that good. However, we do have the first one of the sets in shape.

__Engr:__ Okay, is that the tooth contact frequency at the first location - the one I thought was most important?

__Stat:__ Yes, that's right (handing set of tables to engineer).

__Engr:__ You mean this table here?

__Stat:__ In Table 1, we have the data unshuffled, as you might say, or in the order of the original chart. And also we have the totals there. You will recall we talked about the totals of the rows; for instance, for 25% load, we have a total of 2063; for the 100%, a total of 2057.

__Engr:__ Are these the figures on the last column here?

__Stat:__ Yes, and for the 125%, 2098.

__Engr:__ How can you make any sense out of these data?

__Stat:__ We have a very powerful tool in Statistics called the analysis of variance, whereby we can break down the total variation among all these readings. If you will notice, they range from 61 up to 95.9. Just as you have measure for noise or length, the statistician has a means for variation. So we can break down the total variation and get a measure of the effect (now, that's a new word), that is, the amount of variation that is due to a change in load or volume or speed or direction or combinations of those. I know that is pretty difficult to follow.

__Engr:__ Well, so far it is just words.

__Stat:__ Well, let's take a look at the table and see if that will help out. Look at Table 2. That's the table with direction across the top and speed along the side. Now let's look at the totals for directions. They are 1549, 1555, 1570, and 1543. They are all pretty close together, aren't they?

__Engr:__ Yes. They seem to be.

__Stat:__ Well, we can actually, by a little arithmetic, measure the effect of the differences, that is, how important they actually are. Now, look at the two speeds; 300 rpm is roughly 3200, and 1200 is roughly 3000.

Quite a lot of difference, isn't there? We will also measure the importance or significance of that difference.

Engr: Can you show me physically?

Stat: We have done the arithmetic for this breakdown. It appears in Table 4. If you will notice, the first column of this table are those comparisons we talked about the other day; only we call them degrees of freedom. You don't mind, do you?

Engr: I don't know what they mean anyway, so it doesn't make any difference what you call them.

Stat: Then we have, as I said, broken down the total variation into the part due to each factor and their interactions. First, we got the total variation by finding the average of all the readings, then sub-tracting it from each of the 72 readings, squaring each of these 72 differences, and finally summing the squares. That is the total found at the foot of the second column. By some more arithmetic, we broke this down into the part due to each of the items in the left side head-ing. Next we get the average effect of each one in the last column by dividing each by its corresponding degrees of freedom.

Engr: I hope you don't expect me to get all this.

Stat: Well, not all at once. Now look at the last item in your table, that 15.75. If you took the square root of that, you would have the standard deviation - that is familiar to you, isn't it? - and it would be in decibels, approximately 4.0 decibels.

Engr: I have been reading about standard deviation in journals. It seems that you statisticians are getting in everything these days.

Stat: We're trying to, anyway. We're infiltrating.

Engr: And that is a standard deviation? And that is what is used in normal curves?

Stat: That's right. That is the error term.

Engr: Did you say that was in decibels?

Stat: Decibels, yes.

Engr: So that applies to this test physically? And it is a quantitive measure. It is quantitive.

Stat: That's right. The 15.75 is a measure of the within-day variation or error. Now let's look up the other error term in the main part - that is 8.225 - that measures the day-to-day error.

Engr: Well, that is smaller than the other.

Stat: Only slightly so.

233

Engr: Do these things become more important as they become larger or smaller?

Stat: Oh, as they become larger. The smaller they are, the more sensitive is your test.

Engr: So this means that our day to day variation . . .

Stat: . . . is no greater than the within day variation.

Engr: Well, that is quite interesting because we are spending a lot of time calibrating the equipment every morning; and this means that at least for these two weeks we have the calibration under control?

Stat: Exactly. Now, we use this error term, the 15.75 to decide whether or not a factor in the sub-part is important. If it is no larger than that, then that factor is not important in the test. We'll start at the bottom and work up. The first three, direction by lead, volume and temperature, aren't sufficiently different from the 15.75 to be considered significant. Neither is direction with speed, 8.81 important. Hence, we can conclude that direction does not interact with any one of the other four factors in the test.

Engr: That's really interesting information.

Stat: Now let's look at direction itself . . with a mean square of 7.75.

Engr: I gather that direction is not important either, since it is that small. Well, wait until I get back to the Lab. You know, we had a little argument, friendly, of course, as to this direction factor; and this is evidence that it is not important. This means that we in our test work could just as well only run one direction? Well, good deal! That will cut our test work down to one-fourth.

Stat: That's right. And it is very good evidence too, because it is based on 18 readings of each one of these 4 directions. Now, let's look at S x L. I have prepared a table of speed - load totals, Table 3. If we plotted load along x-axis and decibels on the Y-axis, the totals for 300 rpm would go down a little at 100% and up slightly at 125%.

Engr: Well, that is about flat, isn't it?

Stat: Yes, we find 1200 starts at 964 and goes to 990 and then on up to 1020. Now look at the measure of that effect; that is, lead times speed. We get 68.34, which is quite large. We compare it with our 15.75; that is, we find the ratio of 68.34 to 15.75.

Engr: How do you decide whether it is large enough?

Stat: Well, it has all been worked out mathematically, but I don't think we need to go into its nature here. This ratio of the number to the error term is

Engr: Ratios? Oh, you use ratios?

Stat: Yes, we called these ratios, F ratios. Where "F" happens to

stand for Fisher, the man who is most responsible for this technique.

Engr: You mean he is the one to blame? What do you use, formulas?

Stat: No, tables have been worked out which tell us just how strong the evidence is that a factor is important.

Engr: Oh, you use tables. I see.

Stat: Well, we could go on . . .

Engr: I have an appointment soon, Miss Day. I will tell you, from the looks of this, it seems I can assign the equipment up to some other project because I don't think we have to run this over again.

Stat: Well, I don't think so, either.

Engr: We seem to have enough information; and it is all straight-forward, assuming, of course, the other locations give the same answer.

Stat: We will work up the rest of the data and write you a report on it.

Engr: Well, actually, Miss Day, this has been an experience and I am especially interested in this technique of analysis. That strikes me as being very powerful. You know, I have been avoiding your class in Statistics. Perhaps I should sign up for your class.

Stat: We would be very happy to have you. It has been a pleasure working with you on this.

Engr: And I will wait for your memo.

Stat: Yes.

END OF ACT III

235

Table 1 Test Data (Decibels) From Preliminary Gear Noise Study
3x3 Latin Square Split-plot Design

| Load and Speed | VOLUME | | | | | | | | | | | | | | | | TOTAL |
| | 1/2 gpm Direction | | | | | 1 gpm Direction | | | | | 2 gpm Direction | | | | | |
	TB	IT	IB	TT	Total	TB	IT	IB	TT	Total	TB	IT	IB	TT	Total	
25%	120°					90°					160°					**2063.0**
300 rpm	92.7	95.9	92.7	92.2		91.3	87.7	93.6	92.9		86.9	90.7	92.1	90.6		
1200 rpm	81.4	79.2	85.5	84.4	701.0	68.0	77.7	76.2	72.2	659.6	78.2	97.9	80.2	85.8	702.4	
100%	90° *					160°					120°					**2056.6**
300 rpm	94.2	88.6	89.8	89.8		89.7	87.8	90.4	90.4		91.9	85.4	85.7	82.6		
1200 rpm	80.2	83.7	83.9	75.4	685.6	86.0	86.6	79.3	85.0	695.2	84.8	79.5	86.9	79.0	675.8	
125%	160° *					120°					90°					**2098.1**
300 rpm	88.7	92.1	91.7	93.4		90.3	90.3	90.4	89.7		88.4	86.3	88.3	88.6		
1200 rpm	94.2	91.1	89.2	86.2	726.6	86.7	83.5	86.6	83.0	700.5	75.8	71.2	87.9	84.5	671.0	
TOTAL					2113.2					2055.3					2049.2	6217.7

TEMPERATURE

	90°	120°	160°
	2016.2	2077.3	2124.2

* Missing Data Supplied

‡ Industrial Quality Control, Vol. IX No. 5, March 1953
"The Statistics in a gear-test Program" by B.B. Day and F.R. Del Priore.

236

Table 2. Speed by Direction Totals (decibels)
(Total of 9 observations - loads, volumes and temperatures)

DIRECTION

Speed (rpm)	TB (1)	TT (2)	IB (3)	TT (4)	All Directions
300	814.1	804.8	814.7	810.2	3243.8
1200	735.3	750.4	755.7	732.5	2973.9
Both speeds	1549.4	1555.2	1570.4	1542.7	6217.7

Table 3. Speed by Load Totals (decibels)
(Total of 12 observations - temperature, volumes and directions)

LOAD

Speed (rpm)	25% (1)	100% (2)	125% (3)	All loads (4)
300	1099.3	1066.3	1078.2	3243.8
1200	963.7	990.3	1019.9	2973.9
Both speeds	2063.0	2056.6	2098.1	6217.7

Table 4. The Analysis of Variance of Test Data in Decibels from Preliminary Gear Test
(3x3 Latin Square split-plot design)

Source of variation	Degrees of Freedom (1)	Sum of Squares (2)	Mean Square (3)
Main part (3x3 Latin Square)			
Load, L	2	41.60	20.80
Volume, V	2	103.97	51.98
Temperature, T	2	244.40	122.20
Error (main part)	2	16.43	8.225
Sub-part (split plot)			
Speed, S	1	1011.75	1011.75***
Direction, D	3	23.25	7.75
SXD (1x3)	3	26.43	8.81
SXL (1x2)	2	136.68	68.34*
SXV (1x2)	2	47.81	23.90
SXT (1x2)	2	203.76	101.88**
DXL (3x2)	6	143.89	23.98
DXV (3x2)	6	20.81	3.47
DXT (3x2)	6	146.14	24.36
Remainder (error)	28¢	441.12	15.7542
Total	67	2608.04	

¢ Decreased by four because of missing observations.

* Probability is less than .05 that the differences observed are simply due to random fluctuations.

** Probability is less than .01 that the differences observed are simply due to random fluctuations.

*** Probability is less than .001 that the differences observed are simply due to random fluctuations.

THE QUALITY CONTROL PROGRAM AT
MINIATURE PRECISION BEARINGS, INC.

Charles J. Hudson
Miniature Precision Bearings, Inc.

Ten years ago when the American Society for Quality Control was in its infancy our conversations at meetings centered around intricate detail, many charts, record keeping by machine operators and others and reports issued daily, weekly, and monthly. The mathematics of quality control procedures including probability calculations were emphasized.

In 1958 simplification is the order -- to minimize our record keeping and to retain and circulate only the most important reports. We make our quality control efforts simple and understandable particularly to the machine operators who are the people who can correct plant difficulties causing rejections, reworks and high costs.

What plant exists where it can be truthfully said that the quality control work is completed and nothing remains but routine work? At Miniature Precision Bearings, Inc., as in most other plants, there is still much to be done to correct difficulties resulting in product losses. New problems occur frequently.

The product manufactured at Miniature Precision Bearings, Inc. is instrument and precision ball bearings. About five hundred different kinds of bearings are made, classed as radial, thrust, pivot and angular contact. The outside diameters of these bearings range from 3/8ths inch down to .059". The smallest weighs only .02 grams. They are used in precision instruments and delicate mechanisms such as gyroscopes, recorders, synchros, potentiometers, metering devices, etc. where light weight, long life, accuracy and precision, low torque and a minimum of space and weight are among the requirements.

The problems in the manufacture of miniature ball bearings center around holding dimensions of inner and outer rings to .0002" or less, the sphericity and size of the balls to within a few millionths of an inch and surface finishes of critical surfaces to within 1/2 micro-inch. Concentricities and parallelism of sides or faces are of prime importance. Operating factors such as low torque and low noise requirements are frequent customer specifications.

To obtain close tolerances quality control methods are used for determining machine capabilities. Because parts are small and must be held to narrow limits, machines producing these parts must be held very close to, and sometimes better than, the manufacturer's recommendations; as a result it is necessary to make extensive use of machine capability studies. Frequency distribution charts are often relied upon and sampling inspection by roving inspectors during the manu-

facturing process is prevalent. Roving inspection and operator inspection is followed by a system of toll gate inspection where parts are examined at the end of individual or a series of related manufacturing operations, the "toll" being compliance with specifications.

The problems of maintaining close tolerances in the manufacture of the tiny components entering into small assembled bearings are many. For example, commercial gages for measuring some of the requirements are often not available from the usual source of supply for larger ball bearings, so must be manufactured in the plant instrument shop or purchased as special equipment from gage manufacturers.

The cross sectional dimensions of the outer ring are so small in some cases that the ring can become distorted and an erroneous diameter reading can result from measurements unless contact pressures are kept under two ounces. Amplified electronic or air gage methods are frequently used. Instruments for measuring noise, torque values and radial and axial play frequently must be made. A high percentage of the employees are supplied with microscopes to aid in manufacturing detail.

Dimensions must be carefully controlled to comply with, or surpass, industry standards as established by the Annular Bearings Engineering Committee, to assure proper fits at assembly and to meet customer requirements. Some dimensional tolerances are held to within a few millionths of an inch. In addition surface finishes of critical surfaces such as balls and ball races must be measured and controlled to within narrow tolerances -- 1/2 microinch in most cases.

Methods of maintaining small dimensional tolerances and surface finish requirements are similar to those described in quality control literature. The chief differences are the need of careful control of gage wear, manipulation of gages and adaptation of refined methods. The need of a good metrology section and cooperation with it is of course a requirement.

Miniature ball bearings, like large ones, are composed of three essential parts -- an outer ring, an inner ring and a complement of balls between the two located in the raceways ground into the two rings. Balls of only .015" in diameter are used in some of the bearings.

Manufacture begins with special steel rod, about twelve feet long, fed into standard screw machines of the smallest type. Inner and outer rings are turned with the use of tungsten carbide tools to dimensions approaching the finished requirements in these machines.

Incoming acceptance sampling is performed on the rods and consists of chemical and metallurgical examinations. The samples are examined in the laboratory of the Research and Development section

which works in close cooperation with MPB's Quality Control Department. At the screw machines, samples are taken by the operators and also by the roving inspectors. Performance is examined and controlled principally by simple frequency diagrams.

After hardening, all subsequent metal removing operations are performed with abrasives by grinding, lapping or honing. During these operations, samples are taken by roving inspectors at periodic intervals and examined for dimensional and visual characteristics.

The emphasis on examination is simplicity. The ideal condition is, of course, to have the operator on each machine do his own inspection. The roving inspectors obtain the cooperation of operators and supervision by explaining the nature of any defects.

Easy to read records, cooperation with operators and foremen, early summarization of records showing machine performance, department performance and overall plant performance are some of the steps which make roving inspectors' work effective.

Major steps in manufacture after the screw machine operations are:

> heat treating
> lapping the sides of the rings
> grinding OD and ID of both inner and outer rings
> grinding the ball races in both rings

At each of the above steps sampling inspection takes place at frequent intervals by roving inspectors to make sure that parts are satisfactory to pass along to the next operation. Examinations consist of inspections for:

> dimensions
> proper hardness after heat treating
> concentricities of ball races to outer and inner diameters
> parallelism of ball races to sides or faces of inner and outer rings and of the sides themselves
> surface finishes of the ball races

There are many other inspections such as grooves in the outer ring to hold thin discs only .003" thick which act as shields on the sides of some bearings, proper dimensions of filling slots on full race bearings used to insert the final ball, and many special requirements.

Because the parts are so small and of low weight, transportation through the plant is no problem. Frequent use is made of an examination room, or toll gate, conveniently located. This room is completely equipped with fine gages, frequently checked with master gages,

for the examination of critical dimensions. Product flows through the plant in numbered lots. The lots are brought to the toll gate at frequent intervals where they are subjected to sampling inspection using the Mil. Std. 105A single sampling tables for determining conformance.

Inspection procedures reveal "defects" and "rejections". Defective parts can be corrected by returning them to the proper factory departments. Rejections, of course, cannot be salvaged.

Miniature Precision Bearings, Inc. make bearings to a single high standard of quality known as ABEC-5 grade or better. This procedure assures uniformity of product to all customers.

Having passed the toll gate final inspection, the next step in the manufacturing process is the final assembly operation. When parts reach this department they should be satisfactory and meet all dimensional and geometrical requirements. In addition they have been thoroughly cleaned.

To perform satisfactorily miniature bearings must be absolutely dirt free. Therefore, a great deal of attention and care is taken to maintain the assembly area, the "white room", suitable to assure dirt free bearings. The white room is held to closely controlled temperature and humidity; it is constructed free from obstructions so it can be cleaned easily every day; employees wear lintless clothing; the white room area is kept under positive air pressure with entrance through air locks.

The cleaned parts enter the white room through a single window. Following assembly there is a final inspection, after which the bearings are lubricated with special lubricants, tightly packaged and made ready for shipment.

Oftentimes orders are accompanied with special requirements which can be met best by selective assembly. A common requirement is closer than normal radial play which is usually held to limits of .0002" to .0008" resulting in reasonable clearance for minimum torque and medium speed applications. Low torque requirements, both starting and running, is another example of special specifications.

A program under the supervision of the Quality Control Department is a continuous torque evaluation plan. Torque, ease of rotation, is a property of ball bearings which is desirable to keep low even though not often specified in orders. The program is helpful to engineering and manufacturing in improving a quality characteristic of growing importance.

A vendor evaluation program is also a part of the program. Sampling inspection of incoming products entering into the manufacture of miniature ball bearings is carried on continuously. Reports to

vendors have been mutually helpful with an improvement in vendor products resulting.

Quality Control methods, in the manufacture of tiny products manufactured to close dimensions, are readily adaptable and can be used as readily as in the manufacture of a larger product.
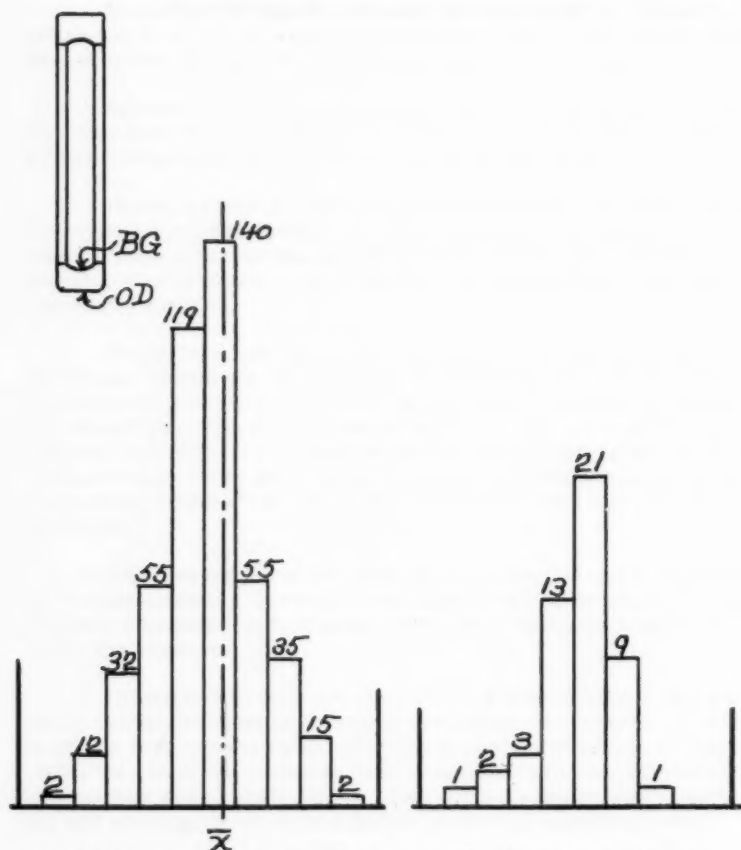
Histogram showing good machine capability for holding ball-groove BG concentric with outside diameter OD of outer ring

Typical periodic capability histogram of screw machine performance

A DESIGNED EXPERIMENT TO EVALUATE SEVEN YARN LUBRICANTS

Hubert M. Hill[a] and Douglas Wheeler[b]

[a]Research Laboratories and [b]Yarn Development Division,
Tennessee Eastman Co., Division of Eastman Kodak Co.,
Kingsport, Tennessee

INTRODUCTION

This experiment was designed to evaluate seven yarn lubricants and two size formulas. The method of evaluation consisted of preparing warp beams with the proper lubricant and size combinations and weaving them into cloth. The cloth was then evaluated for warp defects. The particular design chosen made possible the use of a number of looms for weaving the cloth. It was desired that the experiment show any difference present between the lubricants and allow the best lubricant to be picked from the group.

EXPERIMENTAL DESIGN

The basic design in this experiment was a replicated 2 x 7 factorial design. In addition, a further complication was present in that it was necessary to use several looms in order to finish the weaving in a reasonable time. Since the warp beams were approximately 1,500 yd. long and would require about 15 days of weaving time, weaving one complete replicate on a loom would require nearly seven months. Therefore, in order to obtain a satisfactory design, it was decided that the blocking would be accomplished by the use of a randomized incomplete block design. The combination of factorial designs with randomized incomplete block designs was described by Kramer and Bradley [1, 2]. The particular plan used in this study was S-40 [3], which is shown in Table I. This plan involves 14 varieties replicated three times in seven blocks of six. Kramer and Bradley showed that the 2 x 7 group of the plan can be associated with a 2 x 7 factorial, as indicated in Table II. The experimental design therefore required that three warp beams be prepared from each lubricant-size combination. These were then woven into cloth on a group of seven standard Draper looms, the warps being assigned to looms on the basis of the plan shown in Table I. The warps were run on the looms in random order.

COLLECTION AND ANALYSIS OF DATA

The cloth produced from each warp was cut into approximately 100-yd. sections and evaluated by visual inspection. The defects were rated on the basis of the scale shown in Table III. The total score and the length of each cut were recorded. Data from the first and last cuts were not utilized because previous evidence had shown them to be significantly higher in score.

Table IV shows a typical group of three beams from the same lubricant-size combination $[L_8S_1]$. It should be noted that not only is there a considerable difference in the averages for the three warps which were woven on different looms but there is a difference between cuts in the same warp.

245

Because the scoring scheme is essentially a weighting system for discrete counts, which were expected to be Poisson in nature, it was necessary to transform the data. Experience from previous experiments indicated that either the $\sqrt{x}$ or $\sqrt{x} + \sqrt{x+1}$ transformation should be used. In this experiment it was possible to calculate the mean and variance of the values within each warp beam. These values indicated that on the average the ratio of the standard deviation to the mean was approximately 1.0, which would indicate a log transformation. However, the wide range [0.34 to 1.77] of the ratios casts some doubt on this transformation. Also, the same conclusions were reached by using the log or the $\sqrt{x}$ transformation. Therefore, the $\sqrt{x}$ transformation was used.

The adjustment of the treatment totals to obtain the treatment effects is shown in Table VI. The T_{ij} values are the totals for the three beams for each treatment. The B_{ij} values are the totals for the three blocks in which the treatments occur. The t_{ij} values, which are the adjusted treatment effects, were calculated using the following formula:

$$t_{ij} = \frac{1}{3} T_{ij} + \frac{1}{21} \sum_{j=1}^{2} T_{ij} - \frac{1}{18} B_{ij}. - \frac{1}{126} \sum_{j=1}^{2} B_{ij}.$$

The other values are self-explanatory. A ready check on the calculations is available because $\Sigma t_{ij} = 0$.

From the t_{ij} values an analysis of variance is possible using the information given in Table VII. The analysis of variance as calculated is shown in Table VIII. Although the adjusted treatments do not show any significance, the lubricants are significant at the 90% level.

Returning to Table IV and the distribution of values within a warp, it is believed that some of the higher values are caused by some variable other than those studied in this experiment. If our assumption of a Poisson distribution is correct, it should be possible to determine if any of the values are excessively high when compared with others in the same warp. As a test for high values in a Poisson distribution, we used Molina's Tables of Poisson's Exponential Binomial Limit [4]. The calculated average for the warp was used to check the values against the corresponding Poisson distribution, and any count which had a probability of occurrence of less than 0.001 was excluded. A new average was then calculated and the remaining data were checked against this new average. This test has been used in previous experiments, and in each case except the present one, has resulted in reduction of the variance to the value expected from a Poisson distribution.

By means of this test, approximately 18% of the cuts were removed as having excessive values. These adjusted data were then analyzed. The results of the analysis are shown in Table IX. In this case the lubricants are significant at the 95% level and closely approach the 97.5% level. Once again, neither the sizes nor the lubricant-size interaction are significant.

The treatment averages can be calculated from the grand average and the $\bar{t}_{i}$ values. These averages for the analysis both with and without high counts are shown in Table X. The order of the ranking in the two procedures is essentially the same and the values appear to separate into three groups. However, when the data are compared with the

confidence limit of the difference of two means it is apparent that lubricant 3 is different from lubricants 4 and 7 but is not different from lubricants 1, 2, 5, and 6. In the same manner it can be shown that no distinction can be made between lubricants 1, 2, 4, 5, 6, and 7. These two groups are indicated by the solid vertical lines. This conclusion is substantiated by Duncan's Multiple Range Test [5]. Removal of the high values in the data resulted in an over-all higher significance level without changing the ranking significantly.

One other analysis of importance which was conducted consisted of calculating the percentage of fabric that was removed because of high scores. These data were transformed as the arc sin $\sqrt{x}$. The results of an analysis of these data are shown in Table XI. In this instance the lubricants show no indication of effect, whereas the sizes exceed the 90% level. The lack of any evidence of a relation between the lubricants and the occurrence of high score on cuts within warps indicates that the procedure for exclusion of high values is satisfactory.

CONCLUSIONS

The results of this experiment indicate that lubricant 3 is superior in performance to lubricants 4 and 7, but cannot be distinguished from lubricants 1, 2, 5, and 6. Lubricants 1, 2, 4, 5, 6, and 7 also form a group in which no differences in performance can be demonstrated. There is some evidence that the amount of excessively high scores is related to the size which was used on the warp.

REFERENCES

[1] C. Y. Kramer and R. A. Bradley, Ann. Math. Stat., 28, 349-361 [1957].
[2] C. Y. Kramer and R. A. Bradley, Biometrics, 13, 197-224 [1957].
[3] R. C. Bose, W. H. Clatworthy, and S. S. Shrikhande, "Tables of Partially Balanced Designs With Two Associate Classes," N. Carolina State College, Agr. Exp. Sta., Tech. Bull. No. 107 [1954].
[4] E. C. Molina, "Poisson's Exponential Binomial Limit," D. Van Nostrand Co., Inc., New York, N. Y., 1942.
[5] D. B. Duncan, Biometrics, 11, 1-42 [1955].

Table I
Incomplete Block Design S-40
$v = 14$, $r = 3$, $k = 6$, $b = 7$, $m = 7$, $n = 2$, $\lambda_1 = 3$, $\lambda_2 = 1$

Groups							
1	8						
2	9						
3	10						
4	11						
5	12						
6	13						
7	14						

Plan						
1	8	2	9	4	11
2	9	3	10	5	12
3	10	4	11	6	13
4	11	5	12	7	14
5	12	6	13	1	8
6	13	7	14	2	9
7	14	1	8	3	10

Table II
Association Scheme

L_1S_1 [1]		L_1S_2 [8]	
L_2S_1 [2]		L_2S_2 [9]	
L_3S_1 [3]		L_3S_2 [10]	
L_4S_1 [4]		L_4S_2 [11]	
L_5S_1 [5]		L_5S_2 [12]	
L_6S_1 [6]		L_6S_2 [13]	
L_7S_1 [7]		L_7S_2 [14]	

Table III
Scoring System for Defects

	Points
Major defect	4
Half major	3
Minor defect	2
Half minor	1

Table IV
Scores on Treatment L_6S_1

Warp No. Cut	24		25		26	
	Yd.	Score	Yd.	Score	Yd.	Score
2	130	0	130	0	90	12[a]
3	102	2	-	-	90	88[a]
4	142	0	145	0	93	8
5	108	4	109	2	89	8
6	110	14	119	0	87	40[b]
7	173	26[b]	80	2	70	8
8	86	18[b]	80	3	108	8
9	91	42[a]	136	0	73	7[a]
10	91	8[a]	100	8[a]	93	64[a]
11	90	72[a]	106	0	89	98[a]
Total	1,123	186	1,005	15	882	341
Average per 100 yd.		16.56		1.49		38.66
Adjusted average per 100 yd.		6.30		0.77		9.75

[a] Removed on first test.
[b] Removed on second test.

Table V
Transformed Warp Averages
[All data included]

Block							Block Total
1	3.00 [1]	4.66 [8]	3.65 [2]	2.52 [9]	4.06 [4]	3.00 [11]	20.89
2	2.51 [2]	2.17 [9]	2.67 [3]	1.87 [10]	3.77 [5]	2.98 [12]	15.97
3	5.11 [3]	3.71 [10]	6.04 [4]	9.78 [11]	6.22 [6]	5.11 [13]	35.97
4	3.57 [4]	2.98 [11]	2.47 [5]	2.56 [12]	4.47 [7]	2.91 [14]	18.96
5	2.20 [5]	1.89 [12]	4.07 [6]	2.69 [13]	1.08 [1]	3.27 [8]	15.20
6	1.22 [6]	3.27 [13]	2.80 [7]	3.18 [14]	2.96 [2]	2.24 [9]	15.67
7	4.44 [7]	8.61 [14]	5.25 [1]	2.79 [8]	3.36 [3]	3.27 [10]	27.72
							150.38

Table VI
Adjustment of Data
[All data included]

	T_{ij}		ΣT_{ij}	B_{ij}		ΣB_{ij}	t_{ij}		$t_{i.}$	$\bar{t}_{i.}$
j	1	2		1	2		1	2		
1	9.33	10.72	20.05	63.81	63.81	127.62	-0.4931	-0.0298	-0.5229	-0.2614
2	9.12	6.93	16.05	52.53	52.53	105.06	0.0522	-0.6778	-0.6256	-0.3128
3	11.14	8.85	19.99	79.66	79.66	159.32	-1.0248	-1.7881	-2.8129	-1.4064
4	13.67	15.76	29.43	75.82	75.82	151.64	0.5424	1.2390	1.7814	0.8907
5	8.44	7.43	15.87	50.13	50.13	100.26	-0.0117	-0.3483	-0.3600	-0.1800
6	11.51	11.07	22.58	66.84	66.84	133.68	0.1376	-0.0090	0.1286	0.0643
7	11.71	14.70	26.41	62.35	62.35	124.70	0.7074	1.7040	2.4114	1.2057
$t_{.j}$							-0.0900	0.0900		
$t_{..}$							-0.0128	0.0128		

Table VII
Intra-Block Analysis of Variance

Source of Variation	Degrees of Freedom	Sum of Squares
Treatments [adjusted]	13	$3\Sigma\Sigma t_{ij}^2 - \frac{1}{3}\Sigma t_{i.}^2$
Lubricants [adjusted]	6	$\frac{14}{3}\Sigma \bar{t}_{i.}^2$
Sizes [adjusted]	1	$21\Sigma \bar{t}_{.j}^2$
Lubricants x sizes [adjusted]	6	by subtraction
Blocks [unadjusted]	6	$\frac{1}{6}\Sigma B^2 - \frac{GT^2}{rv}$
Error	22	by subtraction
Total	41	$\Sigma\Sigma\Sigma T_{ijs}^2 - \frac{GT^2}{rv}$

Table VIII
Analysis of Variance
[All data included]

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square [a]
Treatments [adjusted]	13	25.08	1.93
Lubricants [adjusted]	6	20.66	3.44
Sizes [adjusted]	1	0.01	0.01
Lubricants x sizes [adjusted]	6	4.41	0.74
Blocks [unadjusted]	6	59.86	-
Error	22	35.92	1.63
Total	41	120.86	

[a] Significant value at α = 10% — 3.36
     Significant value at α = 5% — 4.16

Table IX
Analysis of Variance
[High cuts removed]

Sources of Variation	Degrees of Freedom	Sum of Squares	Mean Square[a]
Treatments [adjusted]	13	20.55	1.58
Lubricants [adjusted]	6	14.81	2.47
Sizes [adjusted]	1	0.47	0.47
Lubricant x size [adjusted]	6	5.27	0.88
Blocks [Unadjusted]	6	42.68	-
Error	22	18.05	0.82
Total	41	81.28	

[a] Significant value at $\alpha = 5\%$ — 2.09
Significant value at $\alpha = 2.5\%$ — 2.50

Table X
Ranking of Lubricants

All Data			High Scores Removed		
Adjusted Lubricant Average[a]		Adjusted Warp Average [Score/100 Yd.]	Adjusted Lubricant Average[b]		Adjusted Warp Average [Score/100 Yd.]
L_3	2.17	4.71	L_3	1.69	2.86
L_2	3.27	10.69	L_5	2.60	6.76
L_1	3.32	11.02	L_6	2.64	6.97
L_5	3.40	11.56	L_1	2.69	7.78
L_6	3.64	13.25	L_2	2.81	7.90
L_4	4.47	19.98	L_7	3.61	13.03
L_7	4.79	22.94	L_4	3.93	15.44

[a] Variance $[a_i - a_i^1] = 0.70$
98% Confidence limit of difference = 2.10

[b] Variance $[a_i - a_i^1] = 0.35$
98% Confidence limit of difference = 1.48

Table XI
Analysis of Variance
[Percent of fabric removed due to excessive scores]

Source of Variation	Degrees of Freedom	Sum of Squares	Mean Square[a]
Treatments [adjusted]	13	0.606	0.047
Lubricants [adjusted]	6	0.195	0.033
Sizes [adjusted]	1	0.124	0.124
Lubricant x size [adjusted]	6	0.287	0.048
Blocks [unadjusted]	6	0.554	-
Error	22	0.841	0.038
Total	41	2.001	

[a] Significant value at α = 10% — 0.112
    Significant value at α = 5% — 0.163

253

STATISTICAL CONTROL OF A SPINNING LABORATORY

Philip R. Ewald

U.S.D.A. Spinning Laboratory, Knoxville, Tenn.*

Introduction

A spinning laboratory is essentially a yarn mill in minature with most of the mill's attendent control problems. When the mill handles cottons of all types, as is the case for Knoxville Spinning Laboratory, there are introduced into the control problems variables of a nature which require the application of the science of statistics. (1) When the mill is operated as a laboratory, its instrumentation becomes very important; measurements must be carefully made to insure meaningful control. In a spinning laboratory then, we are immediately required to develop an adequate statistical instrumentation. (2)

Fiber Analysis

The first requirement in the processing of cotton yarns is an analysis of the raw material, the cotton fibers. Nature has introduced such variability into the fiber properties of cotton that in the measurement of these properties, the science of statistics becomes every bit as important as the science of instrumentation. Measurement problems resolve essentially into the job of determining mean values and just how good the means are. There are of course two approaches to the problem, first, homogenize the population to the point where only a few measurements are needed, or second, take enough measurements to establish a mean on the population as is. Either procedure could be expensive if carried to an extreme. As it turns out, in normal practice, a combination of the two methods yields the optimum results.

To establish a reference standard, extreme homogenization was carried out on a batch of well known cotton by effectively doubling some 392,000 times in the mixing process. Then, to establish the statistical accuracy of the instruments themselves on the homogenous cotton, replicated measurements were made on 100 samples. The results are given in Table 1. The fineness and immaturity, upper half mean and mean length, and strength and elongation are measured respectively on the Arealometer, Fibrograph and Stelometer developed at the University of Tennessee by Dr. K. L. Hertel and co-workers. Generally, where applicable, the measurement techniques follow the ASTM standard methods. (3) The data in Table 1 is the basis for establishing a control chart for the laboratory, an example of which is shown in Figure 1. Check measurements are made on the standard once each day at a random hour and if the measurement is over the control limits of 2 standard deviations away from the mean, there is an alert. If the descrepancy is repeated, an investigation is made. As a safety measure for those cases where there is a completely unexplained drift outside the control limits, and such cases have happened, the laboratory
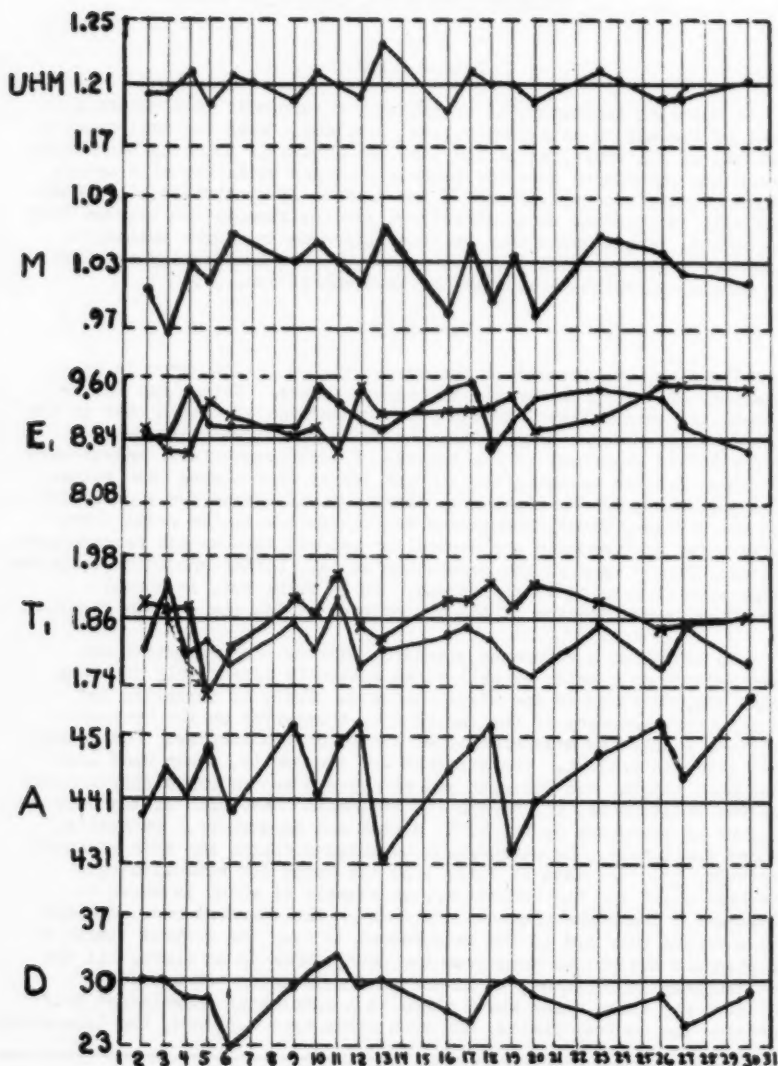

FIGURE 1

uses a correction factor based on the average of the past 10 day's measurements on the standard. Such a procedure of course helps to correct for instrumental drift even though it is within the control limits and no corrective measures are required.

Table 1

Standard K-55 Cotton

Fiber Analysis at 70°F. and 65% R.H.

Values for 100 determinations on completely blended sample.

	Mean	Std. Dev.	CV%	Normal No. Determinations
A Fineness	440	4.8	1.1	2
D Immaturity	26.6	3.6	13.5	2
UHM Length	1.21	.02	1.7	2
M Length	1.03	.03	2.9	2
T_1 Strength	1.88	.06	3.2	4
E_1 Elongation	8.84	.38	4.3	4

As for the routine measurement of fiber properties, an approximation to a homogenized population is obtained by taking 10-15 pinches from the sample under test to give a 5-10 gram sub sample. About 0.5 gram of this sub sample is resampled onto the Fibrograph combs and then homogenized. After length measurements are made, about 0.15 gram samples are taken from the combs for the Arealometer determinations and 0.02 gram samples for the Stelometer. These very small sub samples still contain many hundreds of cotton fibers. The problem next is to determine how many replications are needed in the measurements. Experience has shown that two replications suffice for the length and fineness measurements and four are needed for the strength and elongation. Referring back to Table 1, it is seen that this need follows from the higher coefficients of variation shown by the strength and elongation measurements. The outstanding exception with an abnormally high CV, the immaturity value D, is a calculated value based on the difference of two large numbers which are A values at different pressures, and is accepted for its worth without going to considerable trouble to bring it into line.

With the fiber analyses completed, the data is either referred back to cotton breeders or carried on into yarn processing, depending on the job at hand.

Spinning Process

The fiber analyses are used to set the various operations of the spinning process to optimum efficiency. For example UHM length, and to an extent fineness, governs the speed and settings of the card. Combing is dependent on UHM and M lengths and their ratio, and to an extent on fineness. Drawing, roving and spinning are all critical of UHM length. Roving and spinning are also dependent on fineness. For correlation work, fineness and immaturity are related to neppiness and fiber strength to yarn strength and to some features of processing.

257

In processing cotton fibers into yarns, it is necessary to keep the throughput or weight per unit length of material under process a constant in order to maintain control. Figure 2 illustrates a complete mill process of which the Spinning Laboratory Test is an abbreviated version indicated by the solid lines. Note that feedback

FIGURE 2

is used in each step to keep the process under control. In nearly every case, the feedback consists of the elementary procedure of taking a sample length of material, weighing, and then adjusting the machine if necessary. The effectiveness of this simple control is due to several factors: careful design and stability of the machines, frequent doublings to eliminate random irregularities, and relatively large samples for each measurement. For example, in the spinning test, the samples at the successive stages represent the following percentages: Feed- 100%, card- 100%, draw frame- 11%, roving frame- 25%, spinning frame- 55%. Obviously, where the sample is 100% of the population, the process is under better control than in the cases where the sample is as low as 11%. An improvement can be made if the sampling is done continuously, on all the material, and in a manner which will not disturb the process. Such a sampling can tell much about the short term

variation in the material under process, a factor which is lost in large group samples. Where the operation has been starred in Figure 2, automatic feedback control is already being used in the mills or is being studied.

To make continuous type analyses and study the processing variability, uniformity analyzers have been developed for recording the cross sectional weight of yarns, rovings and slivers. Two approaches to the analysis have been developed. The first type records the weight per unit length of a fixed length of material and gives the mean and variance values for the length. The second type gives a record of weight per unit length and the mean and variance values on a continual basis. It should be noted that the first type will give the same mean and variance values regardless of which direction the material is run, whereas in the second type, the values are affected by previous readings in a continually diminishing degree, depending on the time constant in use. However, it would appear that either method gives about the same amount of information in that the fixed length analysis is just as likely to give different readings for different lengths as is the continual analysis to be affected by past history. A point in favor of the continual analysis is that it may be applied to a process for feedback control.

An analyzer of the second type, using beta radiation which permits 100% sampling on a continuous basis without disturbing the process has been developed at the University of Tennessee. (4) The statistical analysis is attained as shown in the block diagram of Figure 3 which, it will be recognized, contains an adaptation of the machine method for obtaining variance to the continual process. Analysis of this type is
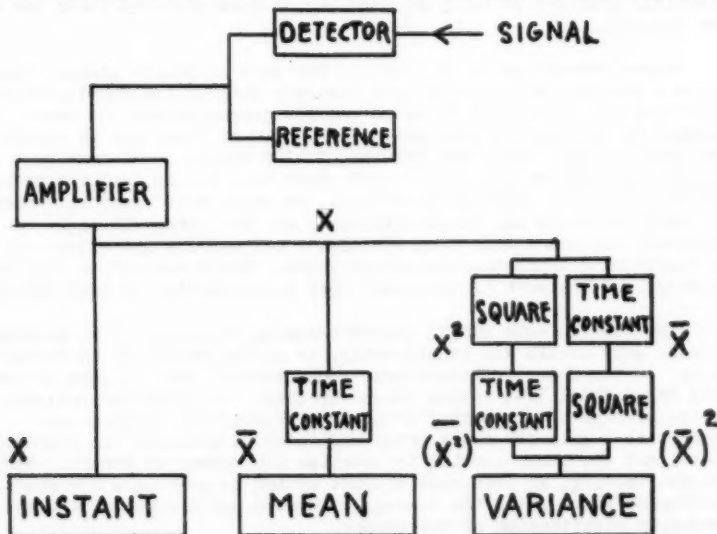

FIGURE 3

259

complete as far as variability is concerned if the irregularities are
normally distributed. In this case, the mean and variance or 1st and
2nd moments completely describe the variability. However, processing
might well introduce 3rd and 4th moment distributions through non-
linearities or instability so that it may well be necessary to start
investigating the automatic analysis of these moments. Periodic
variability is already being automatically analyzed for fixed lengths
of material by an instrument now on the market.

Yarn Testing

In conclusion, the final step of the Spinning Laboratory Test is
the evaluation of the strength of the yarn produced. As might be
suspected, the more yarn that is broken, the better is the average
value obtained. This may be illustrated by studying the variability
of skein and single strand yarn breaking methods used in the Spinning
Laboratory.

The standard method for determining yarn strength at break is to
wind an 80 turn skein using 120 yards per skein which, with 20 breaks
per test, gives a total of 2400 yards used. This method has a CV of
about 3%. In the Spinning Laboratory, a shorter length of about 22.2
yards for a 40 turn skein with 30 breaks gives a total of 666 yards
per test. The CV for this case is about 4.5%. The square root re-
duction in yardage of yarn used is 2 to 1, but the CV has increased
only 1 to 1.5, so it may be said that the smaller skeins are, statisti-
cally speaking, more efficient. Further reduction of the skein length
to 11.1 yards and doubling the turns to 80 at the same time reduced
the CV to about 3.5%, still further improving the efficiency, but the
practical problems of handling such skeins have obviated their use for
the present.

More information is obtained if the yarn is broken single strand,
since a direct indication of yarn strength without the complications
of frictional effects is attained and the determination of yarn
elongation is likewise obtained. However, much less yarn is broken
per machine operation. For example, if 100 breaks of 20 inch gage
length yarn were made, only 55 yards have been tested, and the result-
ing CV of 12%, as found by experiment, is about what might be expected.
The above data was all taken with tests on 22's yarn. It might be
expected that the statistical efficiency per machine operation could
be improved by using heavier weight yarns. Tests have shown this to
be true, and the work of the Laboratory is proceeding in that direction.

In order to make single strand breaking feasible, it is necessary
to make many breaks and consequently, to employ automatic instrumenta-
tion. In the Spinning Laboratory, much research work is done on the
well known Model IP-4 single strand breaker. An automatic indication
of the average of a number of breaks was attained by rigging one
counter to a gridwork and a break indicator to totalize the plane table
travel and the other counter to totalize the number of breaks. For
routine testing, an even more elegant system is used in a commercial
instrument which gives the average of breaks and elongation and also
frequency distribution of the breaks.

Overall control of the Laboratory is attained by frequent com-
parison of the yarn properties with the fiber properties of the cottons

under test and by occasionally running a standard cotton through the
entire process.

Conclusion

In a spinning Laboratory, which is a miniature version of a
staple yarn processing plant complete with a fiber analysis set up,
it is observed that the science of statistics, coupled with an adequate
instrumentation, can result in accurate control of yarn manufacture.

References

1. Applications of Statistics in a Spinning Laboratory. C. B. Land-
        street. Textile Quality Control Papers, Vol. 4, 1957

2. Statistical Instrumentation of a Spinning Laboratory. P. R. Ewald.
        Textile Quality Control Papers, Vol. 4, 1957

3. ASTM Standards on Textile Materials, (D-13 Committee)

4. Uniformity Analysis of Yarns, Rovings and Slivers Using Beta
        Radiation. P. R. Ewald, and C. B. Landstreet. Textile
            Research Journal, Vol. XXVII, No. 6, June, 1957

LIGHTING FOR INSPECTION

Willard Allphin
Sylvania Electric Products Inc.

Note - This paper was prepared primarily for oral
presentation, using slides and shadow-box demon-
strations. The slides are reproduced here as
photographic illustrations, but the demonstrations
can only be described briefly in parenthetical
remarks.

A blind man could work as an inspector if his job were to try
pieces in a go-no-go gauge. However, most inspection tasks involve
seeing, and are therefore wholly dependent upon illumination.

I shall not attempt to give complete prescriptions for lighting
all inspection tasks, but shall attempt to remind you of the prin-
ciples involved and to give some examples of their application.

Lighting engineers like to divide illumination into quantity
and quality considerations, and this will be useful also in discuss-
ing inspection work. We can further divide quality into two categories.
Objective quality, including the direction of light, its color and the
amount of diffusion has an important bearing on the work. So does
subjective quality, involving such reactions as "good or bad,"
"pleasant or unpleasant," "cheerful or depressing." All can affect
the speed and accuracy with which inspection tasks are performed.

Let us first consider quantity of illumination, or more properly,
the brightness of the task. The illumination in footcandles is only
a means to an end. The important thing is the brightness of the task,
since we see things only by the light they emit or reflect.
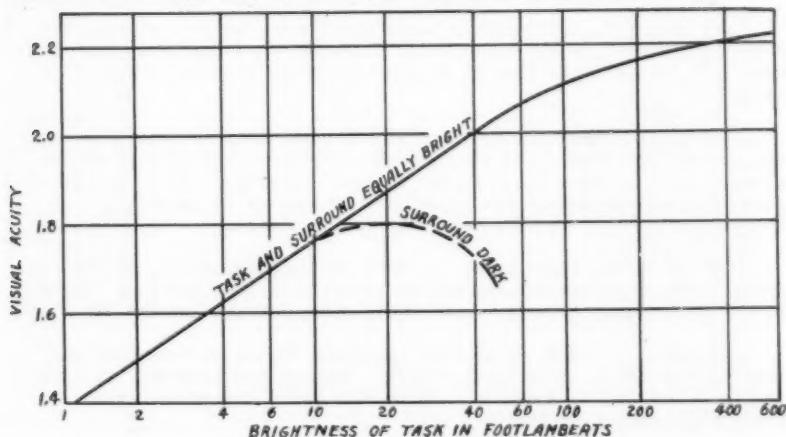

Fig. 1 Acuity increases with the brightness of the
task as long as its surround is bright.

Figure 1, reproduced from the Illuminating Engineering Society Handbook, is typical of how a visual skill can be affected by the brightness of the task. This is for visual acuity, the ability to see small detail, but other visual skills, such as speed of vision, depth perception and contrast sensitivity behave in the same way.

It is true that the dotted curves turn downward after particular brightness levels are reached, but these curves represent cases where the surrounding area was considerably darker than the task. Thus, as the brightness of the task increased, a point was reached where the task itself became glaring enough to interfere with vision. However, glare is a relative thing; and when the surround was bright and increased in brightness with the task, the solid curve shows the result. Acuity continues to increase with brightness even up to sunlight levels of illumination.

Here is a practical demonstration of what I have been saying. In order to reach brightness which will be effective at a considerable distance, I have lighted the task by transmitted light instead of reflected light, but the results could be shown with reflected light if you were close to the task, as you would be in most applications.

At this brightness level I am sure there is a limit to how much detail you can see in the target. (Shadow box with small lighted circular target area in dark surround.) When I increase its brightness you will be able to see more detail and see it more easily. (Target brighter. Surround still dark.) Now if I increase the brightness to this level, it is so high as to be glaring; and you can see less. (Target very bright. Surround dark.)

However, here is the important effect of relative brightness. When I increase the brightness of the surround, you can see better than ever, yet I have not changed the brightness of the target. (Target as before. Surround bright.)

Applying this principle to industrial tasks, we should furnish enough light to make the task easily discernible, and make the surround bright enough so that it will not contrast greatly with the task.

Beyond the immediate surround of the task, quality considerations apply both objectively and subjectively. Fixtures should not produce glare which will make it difficult to see. Ceilings, floors and other significant areas should not be so dark in tone as to cause the eyes to re-adapt every time they look away from the work.

Not precisely measurable, but none the less important, is the "pleasantness" of the environment as affected by the lighting. Color here plays an important part and is, in turn, affected by the lighting.

Direction of light is another important factor in revealing detail, particularly in inspection work. Suppose the inspector is looking for scratches on a surface. With diffused illumination, the surface may look like Figure 2.

Fig. 2   Scratched surface illuminated with diffused
          light.

But when we throw the light in a particular direction, from a glanc-
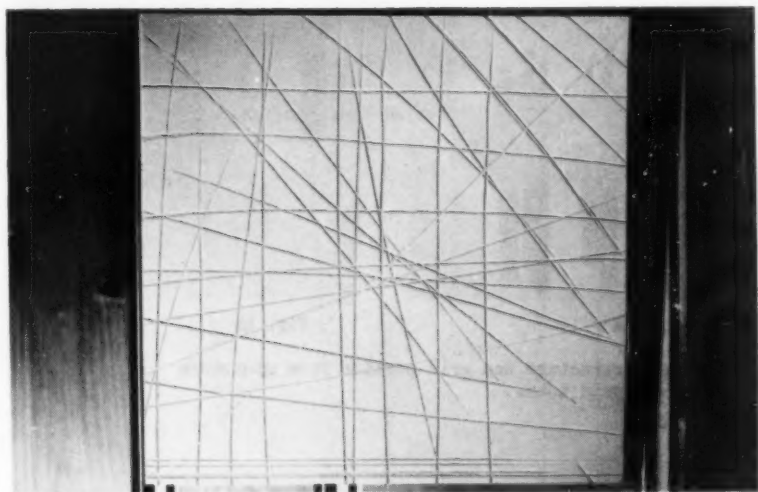ing angle, the surface may look like Figure 3.

Fig. 3   Same surface as in Fig. 2, illuminated with
          "sharp" light from glancing angle.

An extreme case of changing the direction of light occurs when we can
see a particular task better in silhouette. This may be done by plac-
ing the object in front of a lighted area or, in the case of small
parts, an optical comparator may be used.

Figure 4a shows the mount structure of a miniature electronic
tube, and Figure 4b shows the central portion turned 90° and en-
larged further. The wire is considerably smaller in diameter than a
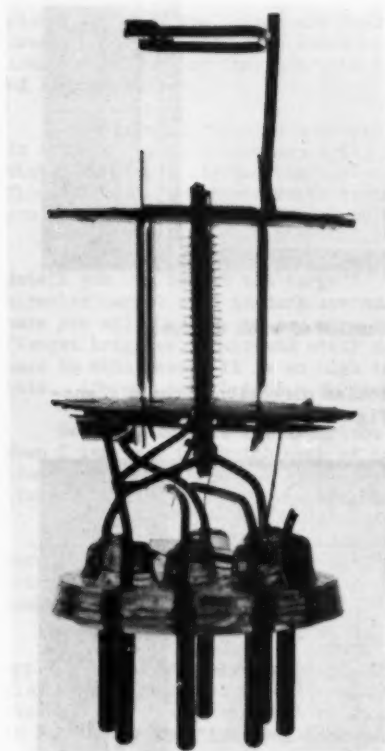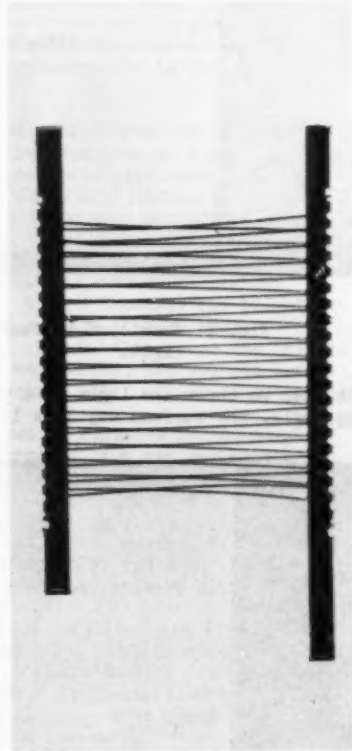human hair.

Fig. 4a   Fig. 4b

Mount structure and grid portion from miniature
electronic tube.

Another type of backlighting is used when cloth is moved across a "perch," as in Figure 5, where the light passes through the product, revealing faults in weaving.
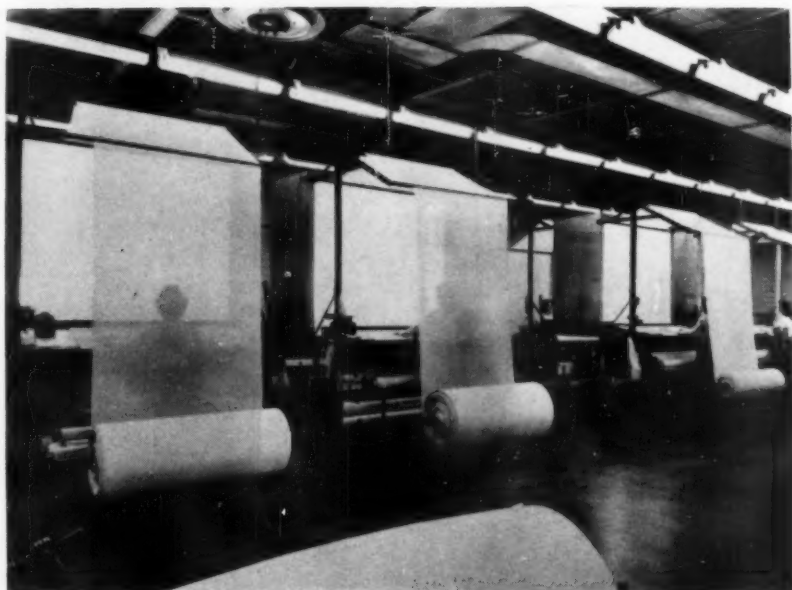

Fig. 5 Inspecting cloth on "perches" by transmitted light.

Edge lighting will often show flaws in glassware, as in Figure 6.

Fig. 6 Inspecting glassware by edge lighting.

A special case is where the material being inspected has a surface which acts like a mirror. If the problem is to see waviness of the surface, this can be shown by reflecting in the surface an image of an illuminated grid, as in Figure 7. The method can be used for inspecting such things as sheet mica, as well as the more obvious shiny sheet metal.
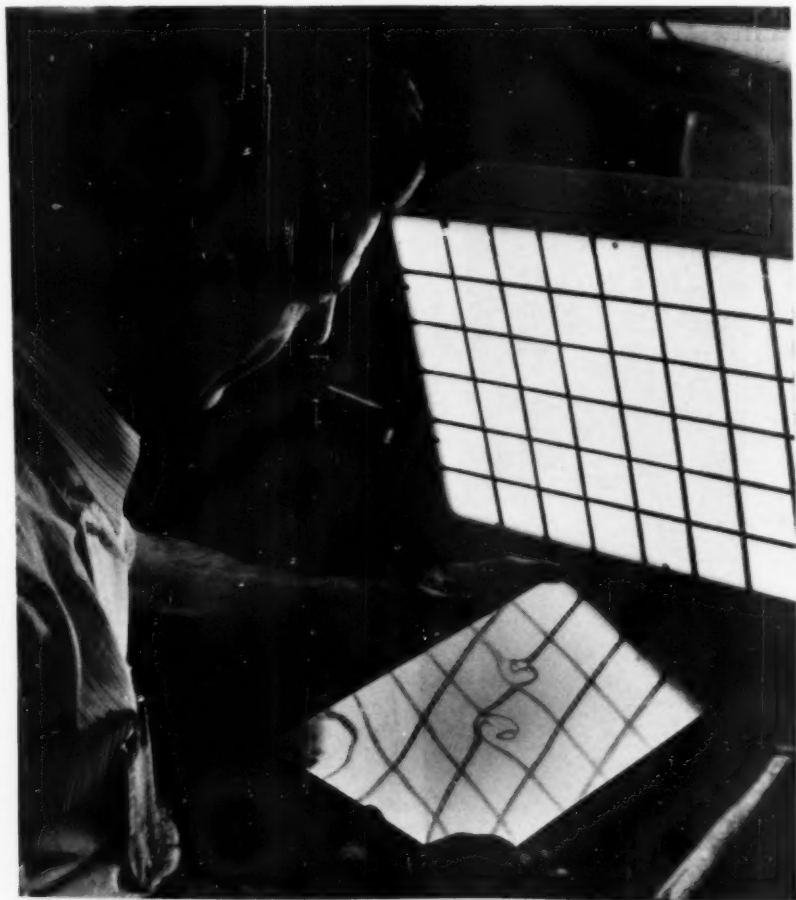

Fig. 7 Inspecting a shiny metal surface by viewing
a reflected pattern.

269

Sometimes reflections must be reduced in order to see details under the gloss of a surface. In such cases polarized light may be the answer.

Some substances have a natural tendency to fluoresce under black-light, that is, ultraviolet radiation just beyond the visible range. If surface flaws cannot be seen properly under visible light, they may reveal themselves under blacklight. A patented variation of this is where a casting is exposed to blacklight after having been dipped in a fluorescent penetrant. Cracks which would otherwise be invisible will show up under these circumstances. Figure 8 shows a flawed casting under normal light and under blacklight, after having been immersed in the penetrant.
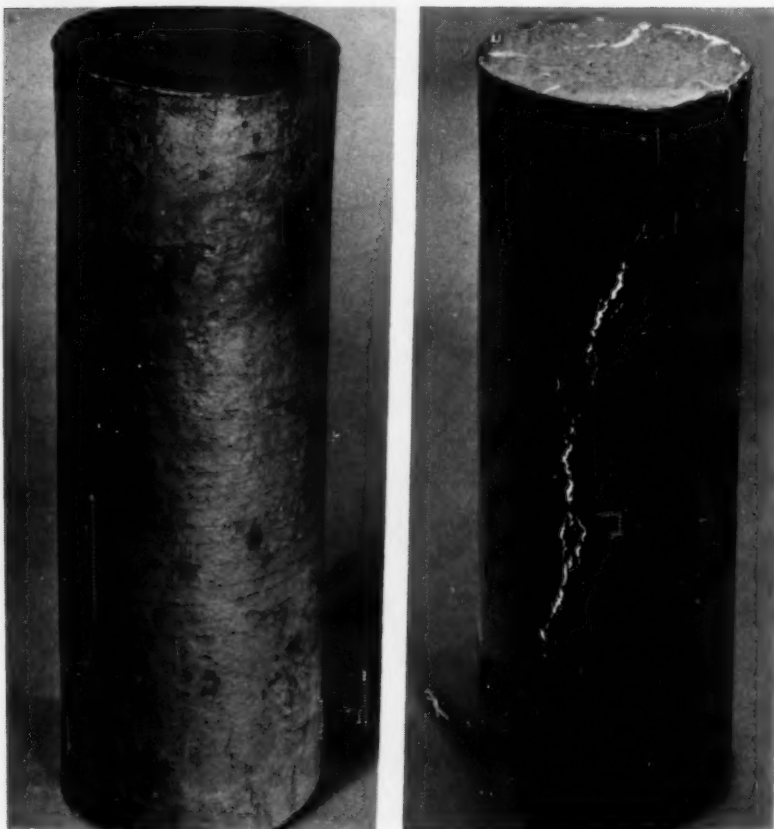

Fig. 8 A flawed casting under normal lighting and
under blacklight after having been treated
with a fluorescent penetrant.
(Photos courtesy Magnaflux Corp.)

270

Except for blacklight, we have as yet said nothing about the color of the light source. In many cases this is unimportant as long as the color is reasonably "natural" in appearance. However, there are times when color is very important. For example, these two cloth samples look much alike, (cloth samples in shadow box) but when I change the color of illumination they appear to be quite different. When two samples behave in this manner they are called a metameric pair.

Sometimes when a complete match is needed between two samples it is necessary to view them under different illuminants in order to make sure that they are not metamers. (Samples which look alike under both colors of illumination.)

Sometimes specifications call for samples to be examined under "daylight," but even where further specified as "north sky daylight" the color will change with time of day and weather conditions. Much greater uniformity is obtained from combinations of fluorescent and incandescent in special "daylight" fixtures. Many times the additional red wavelengths from the incandescent lamps are not really needed. It is more important to have a standardized color of illumination than it is to have a color which exactly matches some particular lighting condition in nature. The human eye is a very adaptable instrument and can readily become used to working under any one of several "white" light sources as long as all of the visible wavelengths are present and the unbalance is not extreme.

Sometimes the balance is purposely upset. For example, a photoengraver finds it difficult to inspect the yellow proof from a color printing plate, because of the paleness of the impression. (Yellow proof in shadow box under white light.)

Illuminated by blue light, the complementary color to yellow, the ink shows up as dark against the white background of the paper, and imperfections in the plate are more readily seen. (Proof illuminated by blue light.)

Inspection lighting should provide more than the bare opportunity to see the required detail; it should enable the inspector to see as easily and comfortably as possible. It is not possible as yet to scientifically measure fatigue, but the common sense viewpoint would certainly indicate that fatigue should be less at the end of the day when the lighting has been as revealing and as pleasant as possible.

There are many types of inspection problems which have not been mentioned here, but each should be met by a joint effort between the quality control engineer and the lighting engineer, the quality man to define what must be seen and the lighting man to recommend the means.

THE USE OF KEY SORT CARDS FOR COLLECTING AND
ANALYZING DATA IN THE MANUFACTURE OF LARGE WIRED EQUIPMENT

George P. Lewett
Western Electric Co., Inc.

Introduction

This article reports on the use of key sort cards for the collec-
tion and evaluation of quality data at the Western Electric Company's
Kearny, New Jersey Works; the report is supplemented by an example of a
control application and a section dealing with the statistical analysis
involved. While the Western Electric Company has, for many years, pio-
neered the development and application of quality control techniques,
further work is being carried out continually. The particular illus-
tration used deals with a type of telephone equipment designated as
wired equipment, in which major operations are the assembly of com-
ponents and their interconnection by the use of a large number of wires.
The control problem centers about the quality of the soldering performed.

Effective control of product quality depends to a great extent on
proper analysis of data. In general, the nature of large wired equip-
ment and the manufacturing processes peculiar to it make analysis par-
ticularly difficult. Information on quality obtained through inspection
and test is difficult to handle and evaluate due to the variety in types
of defects and the many stages involved in manufacture. As a result
therefore, if proper facilities are not provided, the information will
be transformed into an overall measure of shop performance; little use
will be made of details for the specific purpose of isolating and cor-
recting sources of trouble. The Key Sort Card System described in this
article illustrates the type of facility required to permit better use
of the details.

The Manufacturing Process

The process selected for initial application of the Key Sort Card
System is concerned with the assembly and wiring of a telephone switch-
board. The switchboard is one in fairly common use, relatively small in
size (for this class of equipment) and of rather simple circuitry. How-
ever its physical characteristics raise certain manufacturing and quality
problems which are identical in nature to larger wired equipment. It
contains approximately 1000 soldered connections which are made in two
separate wiring stages due to the compact nature of the wiring. Wiring
cycles in the two stages are fairly long so that the production from any
particular operator is numerically small in terms of units per day or per
week. Since it is required to hold insecure connections to a level of 1
in 10,000, a 100% visual inspection of soldering and wiring is made after
each wiring stage followed by operational test and a final inspection of
apparatus mounting, assembly and stamping. For the most part, soldering
defects result from operator error, although defective solder and de-
fective soldering irons are contributing factors. Inspectors perform a
complete inspection of each connection on the unit and are instructed to
report to the shop for repair such defects as excess or insufficient
solder, insecure connections, broken wires, damaged insulation, etc. It
is recognized that differences are to be expected between inspectors in
their judgment as to what constitutes a defect and their effectiveness
in detecting defects. Normal methods of controlling these variables

273

STRAPS: DATE WRD.

QUAN	CODE	507 PBX DEFECT REPORT	REP	CHK
1	41	X5 - B Key #4 - Term 4 Left	✓	✓
1	42	Insecure - T Key #10 - Term 3 Rt.	✓	✓

SURF. WRG.: DATE WRD.

| 1 | 40 | 5X - M Key #6 - term 6 at | ✓ | ✓ |
| 1 | 44 | WNT - ground term | ✓ | ✓ |

TEST: DATE TESTED

1	54	Dirty Contact	✓	✓
1	41	Lamp insulator short - A lamp ckt.	✓	✓
2	51	Relay adjust B & C	✓	✓

CHASSIS   DECK   KEY PANEL

KW 628-Q (3-57)   TICKET NO. D6704

507

Figure 1. Key Sort card used for collecting and
analyzing data on quality (hypothetical
data).

274

are in use but judicious use of the card system can be expected to provide additional means of control.

The Key Sort Card

The card illustrated in figure 1 was designed to tie in directly to the process. The numbered perforations for the most part refer to operators at each of the several stages in the process. The card is attached to the switchboard in the assembly stage and the perforations are marked with a pencil by the various operators, inspectors and testers as the board moves through the manufacturing process. After final inspection the card is removed and punched; the punched perforations then will indicate the route which the particular switchboard followed -- for example: a 507B switchboard, assembled in the first three stages by operators #1, 3 and 7 respectively, strapped by operator #16, inspected by inspector A, etc. The inspectors and testers record their results directly on the face and on the reverse side of the card. The sorting feature is applied by stacking cards from completed switchboards, passing a sorting needle through a particular perforation and lifting the stack; cards punched in that particular perforation will then separate from the main deck. The holes at the lower right, marked "Defects", are to be used on a "special study" basis to give information on particularly troublesome items; those at the top left are used for general control purposes.

100% Inspection and the Need for Good Data Processing Methods

It is an accepted precept of quality control that quality should be built into a product rather than achieved by screening and reworking. Inspection operates according to this principle in two ways: (1) by rejecting poor product for reworking at shop expense, it provides incentive to do the job right the first time and (2) by feeding information back to the making shop it enables the shop to eliminate or control sources of error. In the manufacture of the equipment under study requirements for 100% inspection and the difficulty of processing defect data have combined to hinder the achievement of these two objectives. For 100% inspection to operate effectively in accepting product, two sets of criteria are required: (1) criteria for either accepting the individual product or returning it as a reject to the shop for repair and (2) criteria for accepting or rejecting (halting) the flow of product for sorting and verification. Where wired equipment is concerned however due to the wide variety in types of defects, variations between inspectors, low output of completed units per operator (due to the large number of connections required per unit), etc., it becomes extremely difficult first of all to establish criteria of the second type and secondly to set up means for rapidly comparing operating quality against these criteria. The second purpose of Inspection -- to feed back information -- is also frustrated by the difficulty of processing data so as to reveal sources of defects. The end result is frequently degeneration of inspection to a sorting operation with Inspection becoming responsible for quality.

The Key Sort card is a facility designed primarily to remedy to some degree the difficulties involves in processing data and was stimulated partly by the thoughts expressed in the preceding paragraph. Aside from the sorting feature, the durable construction of the card and the fact that each card gives a complete quality and production history of a specific switchboard are decided advantages over conventional methods. As

275

a trouble-shooting facility however the card evokes the most interest. Trouble-shooting in processes of the type under discussion is basically a matter of comparing two or more equivalent defect sources and determining whether a particular source is significantly different in the frequency with which it emits or causes defects. In order to make valid comparisons however particular consideration must be given to the bases for comparison. In equipment processes this will entail taking into consideration, when comparing for instance the performance of two operators at a particular stage, such factors as type of defect, period of time, tester or inspector identity and operators in preceding or subsequent stages. Due to the numerically low output per operator and the number of levels so to speak of each of the factors listed, this would normally involve handling large quantities of data. The Key Sort card is particularly suited for culling out that data which is pertinent to the problem for on-the-spot diagnosis or for the establishment of preliminary hypotheses. An example of this function is given in a subsequent section.

As a matter of practical application, the completed cards are received initially by the shop supervisor. He is therefore provided with a tool, not heretofore available, for more effectively utilising the information provided by inspectors and testers. Application of the sorting feature enables the supervisor to maintain a continual check on individual operator performance, identify and rapidly run down reasons for major deviations from standard, verify the accuracy and consistency of inspectors, etc.

<u>An Example</u>

Soldering and wiring quality, as measured by the visual inspectors, had been charted over a period of two months in terms of defects per five switchboards (figure 2). The pattern shown however was not completely satisfactory and it was decided to study the results in greater detail. The cards containing the charted results were sorted into all possible combinations of operators and inspectors and each combination was arranged in order of wiring date. Particular combinations were selected
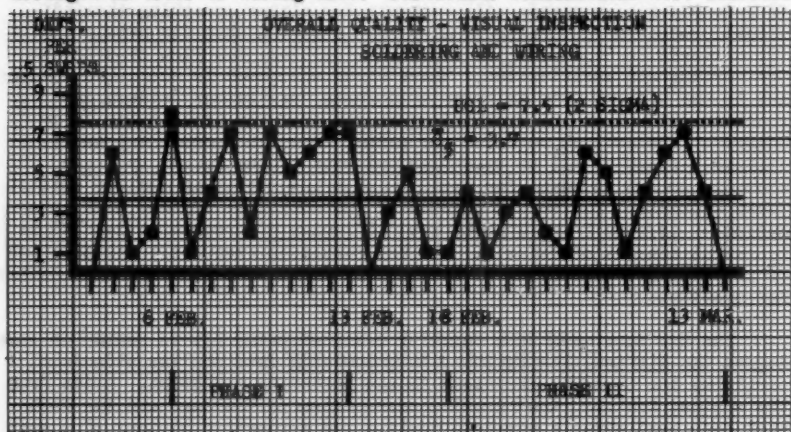

Figure 2. Overall soldering and wiring quality
from visual inspection results.

276

for analysis due to the frequency with which they occurred and their spacing in the time period considered. For each combination selected five successive switchboards were taken starting with February 6 (Phase I) and the results recorded as figure 3A; results on the last five switchboards (Phase II) in each combination were recorded as figure 3B. The data was then analyzed for the purpose of establishing preliminary hypotheses as to probable causes of variation; it was also of interest to determine whether a shift in product quality has occurred and how operator and inspector performance were affecting presentation of group results.

In this example, since analysis was primarily of an exploratory nature, the control chart approach was felt to be particularly suitable. As illustrated in figure 4, the two levels for each factor were first compared in each of the two phases. In these comparisons use was made of the recent report* that the difference between two Poisson variables is distributed with mean $(m_1 - m_2)$ and variance $(m_1 + m_2)$. Hypothesizing no difference between levels, $m = m_1 = m_2$, the expected difference is zero and the standard deviation of the difference is $\sqrt{2m} = s_d$. For purpose of this analysis 3 s_d limits were considered significant although probability levels derived from the exact distribution function were not derived. It can be stated however from consideration of the Tchebycheff inequality that the significance level is at least better than 11%.

For Phase I therefore a center line was drawn at the average for samples of 20 ($\bar{c}_{20} = \dfrac{38}{2} = 19$) and limits drawn at

$$\bar{c}_{20} \pm 1.5\ s_d = \bar{c}_{20} \pm 1.5\ \sqrt{2\bar{c}_{20}}$$

Defects found per 20 switchboards were then plotted for the two levels of each factor -- two points equidistant from the center line $\bar{c}_{20}$. A significant difference is indicated if both points are outside of limits (a difference of 3 s_d). The effect of time was then studied in a similar manner, visually comparing the difference from Phase I to Phase II for each level of the three factors.

The results of the study indicate in brief that there is a strong likelihood of difference between the two inspectors and between the two operators at the Strap Wiring Stage. The general improvement in quality indicated however would seem to be primarily a result of the strong improvement in the performance of operator #4 at the Surface Wiring Stage.

* See reference 1; the frequency function for the difference between two Poisson variables is given as

$$P\ (x) = e^{-(m_1 + m_2)}\ .\ (m_1/m_2)^{x/2}\ .\ I_x\ (2\ \sqrt{m_1\ m_2})$$

where $I_x\ (t)$ is the modified Bessel function of order x and argument t.

3A.

PHASE I   SURFACE WIRING OPERATORS

		3		4		C_{20}
		INSPECTORS		INSPECTORS		
		A	B	A	B	
	1	1	11	4	13	29
			n = 5			
STRAP WIRING OPERATORS	2	1	2	0	6	9
		-15-		-23-		

C_{20} 38 = C_{40}

A - 6 ; B - 32

3B.

PHASE II

		3		4		C_{20}
		A	B	A	B	
	1	3	9	0	3	15
			n = 5			
	2	0	0	0	4	4
		-12-		-7-		

C_{20} 19 = C_{40}

A - 3 ; B - 16

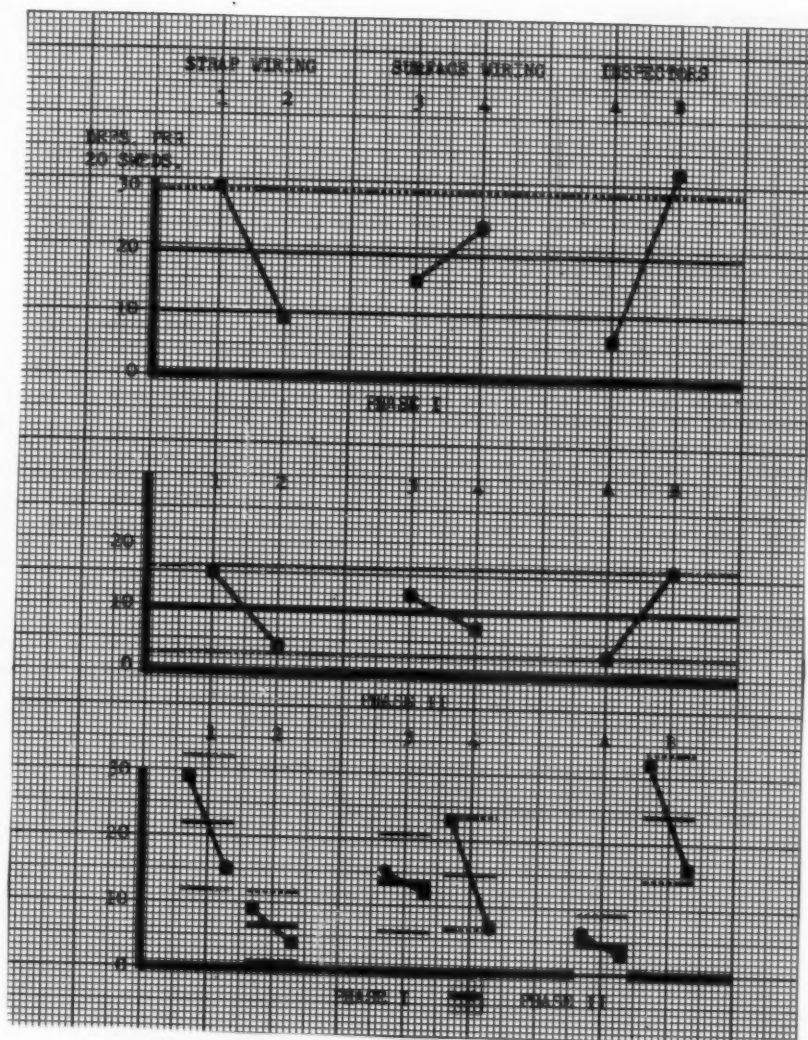Figure 3. Data in defects per five switchboards

278

Figure 4. Analysis of data

Concluding Remarks

Statistical Quality Control is frequently thought of as a collection of specific techniques which have proved to be particularly effective in maintaining product quality at an acceptable level. This concept, emphasizing techniques rather than principles, can be particularly damaging to the cause of Quality Control in some areas of manufacturing. It should be recognized that a particular technique -- such as continuous sampling or use of a standard $\bar{X}$, R chart -- is only applicable under certain specific conditions and in general, it is conceivable that control may be achieved solely through adherence to basic principles. Of these principles, in this sense the most important is the rigorous use of information to control quality at the source. In manufacturing processes, such as that under discussion in this article, adherence to this principle requires development of better methods of collecting and handling information for rapid analysis and transmission to the making shop. The Key Sort card is one such method, of limited application of course, but indicative of the approach needed to effectively install a system for controlling quality in processes of this type. A particular advantage of the Key Sort card is that its durability permits it to travel with the equipment in process and its simplicity engenders rapid evaluation of results for shop control purposes. In addition, it is a permanent record which lends itself to more leisurely statistical evaluation without the need for transposing the data. My remarks have, I hope, illustrated both of these applications.

References

1. "Distribution of the Algebraic Sum of Poisson Variables" ("Distribution de la suma algebraica de variables de Poisson"); Jose Romani, Institute de Investigaciones Estadisticas; Trabajos de Estadistica, 7, No. 2, 175-181, 1956.

2. "A Production Experiment with Mechanical Assemblies"; Ellis R. Ott, Rutgers University; Industrial Quality Control Vol. IX, No. 6, 1953.

THE PROBLEM OF CERTIFICATION THAT MATERIALS
MEET STATISTICAL SPECIFICATIONS

R. D. Smith
Union Carbide Nuclear Company

For a number of years the Y-12 Plant of Union Carbide Nuclear Company in Oak Ridge, Tennessee has been confronted with a materials quality assurance and certification problem which is, if not unique, at least unusual. Our customers have presented us with this problem in the form of specifications for the chemical constituencies of our products. These specifications are written on a statistical basis. In the absence of specific references to similar problems in the quality control literature we have devised several methods for certifying that our products meet these specifications. Since we feel that specifications of this type will probably become more common in the future, we trust that the following brief description of our methods will have some value for others.

To exemplify the type of specifications to which I refer let me invent a chemical compound, say "unium carbide", which, we will say, my plant is committed to produce in considerable quantity over an extended period. For such a compound we might be provided the following specifications:

Chemical Specifications

1. This material shall be unium carbide.

2. Chemical analyses for unium, carbon and impurities listed below shall be performed on as many samples as necessary to fulfill the requirements of the following paragraphs.

3. Using standard statistical techniques, the manufacturer shall be able to demonstrate:

   a. With 95.0% confidence that 97.5% of the material delivered contains at least 0.780 grams of unium per gram of material.

   b. With 95.0% confidence that 97.5% of the material delivered contains at least 0.190 grams of carbon per gram of material.

   c. With 95.0% confidence that 97.5% of the material delivered contains no more than 500 parts per million by weight of iron or sulfur.

4. The manufacturer shall submit quarterly reports giving the results of the chemical measurements and the statistical calculations which demonstrate that the conditions of paragraph 3 have been satisfied.

Aside from a certain ambiguity in the phrase "standard statistical techniques" the specification at first glance seems to be easily followed assuming, of course, that the product is truly acceptable. Serious difficulties can arise, however, in demonstrating statistically that these specifications have been met even though the product is good.

Basically, we have used four distinct statistical techniques for this purpose. The first and simplest of these we call the total variability method. Suppose we consider the impurity sulfur for which the specification set an upper limit of 500 parts per million in our imaginary compound. Let us say that our analyses for sulfur average 50 parts per million and that these analyses have a variance of 2500 parts per million squared. Then use of the standard deviation in conjunction with the tolerance factor, $K^{(1)}$, will provide a certification statement which is satisfactory. It should be noted that the variance of the analyses was used directly without regard to the sources of variance. The following table summarizes the above calculations and gives the certification statement:

Number of Analyses	N	20
Mean of Analyses	$\bar{X}$	50
Variance of Analyses	S_t^2	2500
Tolerance Factor	$K_{(N, .95, .95)}$	2.752
Computation	$\bar{X} + K_{(N, .95, .95)} S_t$	188

Certification: We are 95.0% confident that 97.5% of the material delivered contains no more than 188 parts per million by weight of sulfur.

We have met the specification in all respects. We must recognize, however, that the certification statement paints a conservative, rather than an accurate, picture of our out-going product's quality. This is because the variance used includes a component from measurement variability. It has been our experience that chemical analytical measurement methods frequently give rise to variabilities as large as, or larger than, the inherent variability of the characteristic measured. Nevertheless, this total variability method is used whenever possible because of its simplicity. Unfortunately, it is seldom that we are able to operate so far within specifications as to permit use of this method.

The variance of analytical results obtained on a product will invariably be composed of two components. These are (1) the inherent variance of the characteristic of the product to be measured, that is, the process variance, and (2) the variance of the measurement method itself. We term these variance components process variance, S_p^2, and measurement variance, S_m^2. Then $S_p^2 + S_m^2 = S_t^2$. I have already mentioned that, more frequently than not, it is impossible to meet our specifications using the total variance, S_t^2.

The other three certification methods to be described are therefore based on the separation of these variance components. That is, they use only the process variance which, after all, is really the only portion of the variance with which the customer is properly concerned. The certification method which has been used most frequently is the variance subtraction method.

Briefly, the procedure followed is this. The total variance of a set of analytical results is easily obtained. The measurement variance, obtained from a laboratory quality control program, is subtracted from this total variance to obtain an estimate of process variance. The degrees of freedom of this process variance is obtained by Satterthwaite's approximation method.[2] The process standard deviation, together with the tolerance factor, then provides us with a certification limit. Let us take an example using our fictitious element unium:

Number of Analyses	N_1	41
Mean of Analyses	$\bar{X}$	0.7850
Total Variance	S_t^2	.000004
Degrees of Freedom of Measurement Variance	N_2-1	20
Measurement Variance	S_m^2	.000001
Process Variance	$S_t^2 - S_m^2 = S_p^2$	.000003
Degrees of Freedom of Process Variance	$\dfrac{(S_p^2)^2}{\dfrac{(S_t^2)^2}{N_1-1} + \dfrac{(S_m^2)^2}{N_2-1}} = N_3-1$ 20	
Tolerance Factor	$K_{(N_3, .95, .95)}$	2.752
Certification Computation	$\bar{X} - K_{(N_3, .95, .95)} S_p$	0.7802

Certification: We are 95.0% confident that 97.5% of the material delivered contains at least 0.7802 grams of unium per gram of material.

Thus we have a satisfactory certification of the product as acceptable using the variance subtraction method. It will be noted in the example that S_t^2 is four times S_m^2. Suppose the measurement variance component is relatively larger in the example. We then might have:

Number of Analyses	N_1	41
Mean of Analyses	$\bar{X}$	0.7850
Total Variance	S_t^2	.000004
Degrees of Freedom of Measurement Variance	N_2-1	20
Measurement Variance	S_m^2	.000003
Process Variance	$S_t^2 - S_m^2 = S_p^2$	.000001
Degrees of Freedom of Process Variance	$\dfrac{(S_p^2)^2}{\dfrac{(S_t^2)^2}{N_1-1} + \dfrac{(S_m^2)^2}{N_2-1}} = N_3-1$	1
Tolerance Factor		Approaches infinity.

Therefore in this variance situation no certification is possible using this method. With a little thought this result is understandable. The smaller the process variance to be estimated, the larger is our relative uncertainty of the estimate. In the mathematics this is indicated by the diminishing degrees of freedom.

We have now had a look at the second certification procedure used by Y-12, the variance subtraction method. This is a useful tool in many applications but it fails, even on good product, when the measurement variance is large relative to the process variance.

Our third method of certification we call the Chi-square method. This method has at times provided satisfactory certification statements for product which could not be certified by the variance subtraction method. The Chi-square method consists of placing an upper 95 percent confidence limit on the total variance and a lower 95 percent confidence limit on the measurement variance and subtracting these limits.[3] The resulting difference is virtually an upper 95 percent confidence limit on the process variance. The standard deviation obtained from this process variance limit can be utilized together with the normal distribution statistic, 1.96, to provide a certification value that meets the requirements of the specification. Consider, as an example, the situation last described in which the measurement variance was three-quarters of the total variance.

Number of Analyses	N_1	41
Mean of Analyses	$\bar{X}$	0.7850
Total Variance	S_t^2	.000004
Degrees of Freedom of Measurement Variance	N_2-1	20
Measurement Variance	S_m^2	.000003
Upper Limit on Total Variance	UL S_t^2	.00000593
Lower Limit on Measurement Variance	LL S_m^2	.00000144
Upper Limit on Process Variance	UL S_p^2	.00000449
Certification Computation	$\bar{X}$ - 1.96 UL S_p	0.7809

Certification: We are 95.0% confident that 97.5% of the material delivered contains at least 0.7809 grams of unium per gram of material.

As you see the Chi-square method results in a legitimate certification statement using these data where the variance subtraction method failed. For this reason we have often used the Chi-square method. Nevertheless, this method, too, has built-in draw-backs. The principal difficulty is that in spite of increasing degrees of freedom the reliability of an estimate of variance does not increase very fast. Thus we have often found that the limits on our variances were too wide resulting in an upper limit on process variance which was too large to permit certification of a product. You will note in the above example that the upper limit on the process variance is larger than the total variance with which we started. A few less measurements and certification would have failed.

The methods I have described thus far are those that occurred to us immediately for meeting the specifications. I have mentioned the difficulties encountered in using each of these methods. Fortunately, we never failed to be able to certify a product we believed to be actually acceptable using these methods. But it was recognized that such a failure was possible.

Another consideration bears mentioning in connection with these methods. Although a mechanic were to select the wrong socket wrenches several times before finding the right one to turn a certain nut he would certainly not be accused of a lack of ethics. But, as you may have noticed, laymen tend to distrust statisticians. These certification methods have the peculiarity that if one procedure certifies the product, the product is good even though the other procedures failed. As in the case of the wrong socket wrenches the difficulty is in the tool used. Still, the statistician is open to the criticism that in order to certify sub-standard product he juggles certification methods

285

until he finds one that will do the job. Obviously it would be desirable to have a universally applicable certification method which, like a monkey-wrench, would fit any problem presented.

We are indebted to Mr. A. de la Garza[4] of our company for developing what we believe to be such a method. The derivation of this method is as follows:

Population variances exist such that

(1) $\sigma_p^2 = \sigma_t^2 - \sigma_m^2,$

σ_t^2 is the total variance with degrees of freedom, N_t,

σ_m^2 is the measurement variance with degrees of freedom, N_m, and

σ_p^2 is the process variance.

Select a function ϕ such that

(2) Prob. $(\sigma_p^2 \leq \phi S_t^2 - S_m^2) = 1 - \alpha$

in which α is small.

Then

(3) $\dfrac{S_m^2 + \sigma_p^2}{S_t^2} \leq \phi$

(4) $\dfrac{\sigma_m^2 \left(\dfrac{S_m^2}{\sigma_m^2} + \dfrac{\sigma_p^2}{\sigma_m^2} \right)}{\sigma_t^2 \dfrac{S_t^2}{\sigma_t^2}} \leq \phi$

Observe that

(5) $\dfrac{S_m^2}{\sigma_m^2}$ is distributed as $\dfrac{\chi_m^2}{N_m}$

(6) $\dfrac{S_t^2}{\sigma_t^2}$ is distributed as $\dfrac{\chi_t^2}{N_t}$

Let

(7) $K = \dfrac{\sigma_t^2}{\sigma_m^2}$

Then

(8) $K-1 = \dfrac{\sigma_p^2}{\sigma_m^2}$

286

Substituting in (4) above, we have

$$(9) \quad \left[\frac{\frac{\chi_m^2}{N_m} + (K - 1)}{K}\right] \Bigg/ \frac{\chi_t^2}{N_t} \leq \phi$$

It will be seem that the expected value of the numerator of this expression is as follows:

$$(10) \quad E\left[\frac{\frac{\chi_m^2}{N_m} + (K - 1)}{K}\right] = \frac{1 + (K - 1)}{K} = 1$$

and the variance of the numerator is

$$(11) \quad V\left[\frac{\frac{\chi_m^2}{N_m} + (K - 1)}{K}\right] = \frac{V\left(\frac{\chi_m^2}{N_m}\right)}{K^2} = \frac{1}{K^2 N_m}$$

Thus the numerator of the experssion is approximated by the distribution χ_m^2 / N_m with degrees of freedom $K^2 N_m$.

From (9) we now have

$$(12) \quad \frac{\chi_m^2/K^2 N_m}{\chi_t^2/N_t} \leq \phi$$

This we recognize as Fisher's F so $\phi = F (K^2 N_m; N_t; \alpha)$ and from (2)

$$(13) \quad \text{Prob.} \left[\frac{\chi_m^2/K^2 N_m}{\chi_t^2/N_t} \leq F (K^2 N_m; N_t; \alpha)\right] = 1 - \alpha$$

$$(14) \quad \text{Prob.} \left[\sigma_p^2 \leq F (K^2 N_m; N_t; \alpha) \; s_t^2 - s_m^2\right] = 1 - \alpha$$

$$(15) \quad \text{Prob.} \left[\sigma_p \leq \sqrt{F (K^2 N_m; N_t; \alpha)} \; s_t^2 - s_m^2\right] = 1 - \alpha$$

In this expression σ_p represents an upper 95 percent confidence limit on the process standard deviation which can be used with the normal distribution statistic, 1.96, to provide a certification.

287

Let us now use this method with the last set of data:

Mean of Analyses	$\bar{X}$	0.7850
Degrees of Freedom of Analyses	N_t	40
Total Variance	S_t^2	.000004
Degrees of Freedom of Measurement Variance	N_m	20
Measurement Variance	S_m^2	.000003
$F \ (K^2 N_m; \ N_t; \ .05)$	-	1.72
Process Variance Limit	σ_p^2	.00000388
Certification Computation	$\bar{X} - 1.96 \ \sigma_p$	0.7811

Certification: We are 95.0% confident that 97.5% of the material delivered contains at least 0.7811 grams of unium per gram of material.

It will be noted that this certification is in good agreement with that obtained using the Chi-square method which resulted in a certification level of 0.7809. An interesting variation in this method is possible if the ratio of total variance to measurement variance is close to one. This ratio, if you remember, is K which appears as $K^2 N_m$, one of the degrees of freedom used in looking up the F value. If K is nearly one, the K^2 is also nearly one, and may be neglected which simplifies the calculations. In the case above if K^2 is neglected the certification level becomes 0.7809. This demonstrates that the K ratio does not have to be very close to one to make neglecting it possible. When K is neglected our certification statement can read "We are <u>at least</u> 95.0 percent confident, etc".

Insofar as we have been able to ascertain this procedure using Fisher's F will certify as acceptable any truly acceptable product if a reasonable amount of data has been collected for the variance calculations. The broad applicability of this method will eliminate any suspicion that the statistician adapts his certification methods to the product.

References

1. Eisenhart, C., Hastay, M. W., and Wallis, W. A., "Techniques of Statistical Analysis", McGraw-Hill, Ed. 1947, Tolerance Factors for Normal Distribution (Table 2.1) pp 102-107.

2. Satterthwaite, F. E., "An Approximate Distribution of Estimates of Variance Components", Biometrics Bull. Vol. 2, No. 6 (December, 1946).

3. Hald, A., "Statistical Tables and Formulas", Wiley, Ed. 1952, Table IV, pp 44-46.

4. Personal communication.

SQC CAN BE MADE TO WORK

Thomas A. Budne
Rath & Strong, Incorporated

Quality requirements on manufactured goods in almost all industries have steadily increased over the past years. As the supply surpasses the demand for products in some industries, competition creates strong pressures for better quality and for lower manufacturing costs. The current pressures for more reliable military products are well-known to all of us. A science of Quality Control has evolved and grown at a speed, at least sufficient, to meet all these challenges of higher quality standards, lower production costs, and more reliable product. Stimulated by the thinking of Walter Shewhart in his memorable book, "Economic Control of Quality of Manufactured Product" and by the new sampling concepts of Dodge and Romig, the control chart and the acceptance sampling plans became the nucleus for the new science. Tremendous technological progress in measurement equipment, the establishment of basic principles with regard to the quality control function and its place in industrial organization, and a gradual absorption of more and more of the existing and growing body of modern statistical principles and methodology have brought the science of Quality Control to its present stage of evolution.

The tremendous contribution of the Quality Control science to modern industry is quite generally acknowledged and accepted. Most of us who have entered the Quality Control engineering profession can point with pride to one or more particular applications of SQC principles and techniques which have resulted in many tangible benefits to our companies. We look with admiration at a few companies which have been using this new science to its full potential and have been getting exceedingly profitable results.

It appears strange, then, that when asked to recommend a number of companies which might serve as examples of Quality Control progress, we find it difficult to add many names to a relatively short basic list. It would be interesting to note how many of us would add the names of the companies in which we have Quality Control responsibilities. The explanation must lie in the fact that while there has been some QC progress in most companies, it is quite small compared to our estimates of what it could or should be. The obvious question is, "WHY?". If companies can make a better quality product and reduce their quality losses by setting up a good quality control program, why haven't most companies gone all out?

A long list of reasons can be offered, but standing out sharply at the top is:

THE FAILURE TO ACQUIRE TOP MANAGEMENT INTEREST, UNDERSTANDING, APPRECIATION AND SUPPORT and, or because of, THE FAILURE TO PLAN A REALLY PRACTICAL QUALITY CONTROL PROGRAM WHICH CONVINCINGLY ESTABLISHED THE FACT THAT EVERY DOLLAR SPENT ON QUALITY CONTROL WOULD BRING ABOUT A SUBSTANTIAL RETURN IN THE REDUCTION OF EXISTING QUALITY LOSSES.

If we define QUALITY LOSSES as all the costs and losses which would disappear if all quality problems were eliminated, we necessarily include Procurement, Design, Engineering, Manufacturing and all activities on through to sales and customer acceptance investigations. Better quality, then, is unavoidable if Quality Losses are reduced. We shall limit our discussion here to a program for the reduction of quality losses in manufacturing.

Experience indicates time and again that while most measured effects are actually cumulative effects of many causes, a small minority contribute to a major share of the total effect whereas the large majority of causes are responsible for a very small share of the total effect.

When all the losses due to sub-standard product in a manufacturing operation are broken down by products, we generally find that only a few out of many products are responsible for more than half the losses.

When all the variation in a product can be apportioned to each of the sources of variation, we find that a very few of the sources contribute most of the variation.

If all the scrap in a machine shop were broken down by machine, we would generally find that a few of the many machines or one type of machine out of many types are responsible for most of the scrap loss.

This basic principle is essential in planning a QC program in which every dollar spent must account for several dollars of reduction in quality losses. We shall refer to it as the principle of mal-distribution. It clearly indicates that profit-making SQC activity cannot be spread over a whole manufacturing operation like a coat of paint, since quality losses are certainly not spread out that way. This mal-distribution can at times be so severe that as few as 10% of all the categories of quality losses may account for as much as 75% of the total loss. In one particular plastics film extrusion operation, one category of loss accounted for more than 50% of all the losses. An investment of QC dollars could - and did - bring a tremendous return.

If the mal-distribution principle seems at times to be out of operation in getting a distribution of the contributions to a particular effect, it may be that the best categories of breakdown have not been selected. For example, in an iron foundry, a breakdown of total scrap by types of defect indicated an almost equal contribution of scrap by each type of defect.

However, this was a job-shop type of operation involving short runs on several hundred different patterns, with about 100 patterns in each week. A breakdown by patterns showed that 15% of all patterns accounted for more than 2/3 of all the scrap. A QC program which succeeded in eliminating the one most prevalent type of defect could mean a total reduction in scrap loss which was less than the cost of achieving and maintaining this reduction. However, a QC program which succeeded in reducing the scrap for the one largest scrap-producing pattern each week could bring a substantially greater return. As a matter of fact, QC program which placed its major emphasis on the 15 largest scrap-producing patterns each week and on keeping new patterns out of this category succeeded in reducing an average monthly scrap loss of $20,000 to approximately $13,000 in 4 months and to less than $10,000 in 8 months.

Charts previously kept on a large number of measurable operating conditions in sand and metal had been ineffective in accounting for scrap. Their cost was extremely difficult to justify. The QC program led to the fact that only a few of the many controlled operating variables were really important in affecting scrap, and tight control of these few could be maintained at considerably less cost than previously expended on all the operating variables. The mal-distribution principle was in operation. A substantial reduction in inspection hours and in finishing hours and the value of over-all improved quality were not included in the scrap loss figures.

Having selected the areas for profitable investment of QC dollars, a systematic plan of action is necessary. Directly or indirectly, most quality losses are a result of variation in a particular quality characteristic of the product. This variation may manifest itself in an attribute "good or bad" property or in a measurable property which varies too much to remain within specified limits for acceptable product. We shall assume that the importance and validity of the specification and the ability to adequately measure the product against this specification have been established as a very important but nevertheless routine function of the QC system.

Since we recognize variation as a cumulated effect of all non-constant conditions in the manufacturing operation, we again seek to isolate the few non-constant conditions which we expect will account for most of the total variation. We are not particularly interested in noting whether these are assignable or unassignable causes of variation. We are interested only in identifying them as the major sources of variation. In other words if the process capability accounts for almost all the variation, we shall seek those conditions which most affect the process capability. If the process capability is a small part of the total variation, we shall look for factors which would normally be called assignable causes.

If repeated tests on the same piece of electronic equipment account for a large share of the variation, then causes of instability in the test method, the test gear, the test conditions, or in the equipment itself must be found.

In a processing operation, the minute-to-minute differences may be small but the hour-to-hour fluctuations may account for most of the product variation. Factors which can possibly change from hour-to-hour would be sought.

It is possible for us to estimate the portion of the total variation which is attributable to changes occurring within a unit or from unit-to-unit, or lot-to-lot of raw material, machine-to-machine, hour-to-hour, operator-to-operator, shift-to-shift, day-to-day, or any other reasonable category. Each manufacturing operation has categories which may be characteristics of itself. A foundry operation would add: cavity-to-cavity differences, cope-to-drag differences, mold-to-mold differences, batch-to-batch differences in sand, heat-to-heat differences in metal, etc.

Small samples, selected at the appropriate times and treated graphically will supply the information we want. A control chart procedure can be adapted for this purpose, but the multi-vari charts (I) are even simpler and more appropriate for this application.
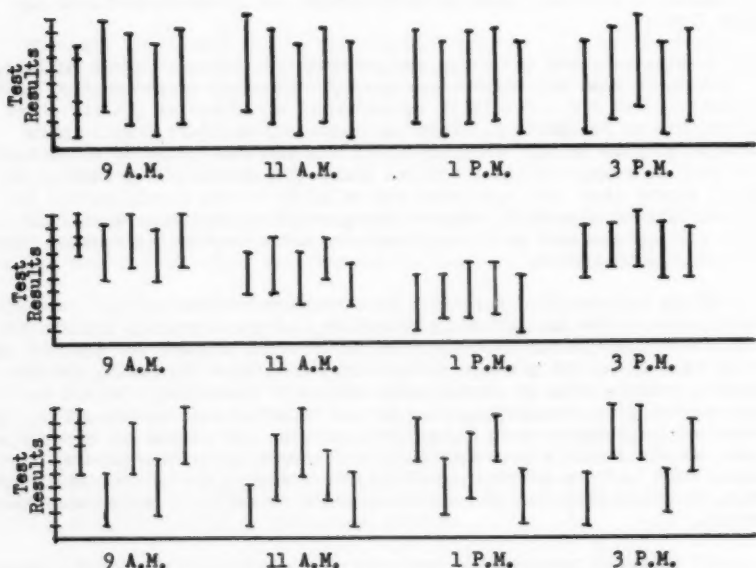
MULTI-VARI ANALYSIS

Figure I

In the above illustration in Figure I, 5 measurements on each of 5 consecutive pieces at every two-hour interval indicate alternative conditions.

The first shows large within unit variation as the major component of variation, the second shows time-to-time variation as the major component of variation, and the third shows unit-to-unit variation as the major contributor.

At this point, experience with the operation may be sufficient to identify the non-constant condition which explains the picture given by the multi-vari charts. A simple one-factor experiment with the suspect condition set at two or three levels will provide the objective evidence required. If two or three alternative conditions are strong suspects, their effects may simultaneously be evaluated by simple experimental designs. If the correct factor has been isolated, the experiment measures its effect quantitatively.

The control chart, the multi-vari chart, or some similar technique provide a systematic first step. Experience, horse sense, and simple designed experiments may sometimes suffice to finish the job. More often than not, however, the easily identifiable trouble sources have long since been isolated and kept in check. There may now be as many theories as there are variables; perhaps 15 to 20 of them including their interactions, which could explain the major category of variation identified in the multi-vari charts. The usual designed experiments fail us by becoming impracticably large and complex. Ten factors at 2 levels of each require 1024 test units, an impossible situation.

The recent development of the Random Balance Experiment (2) by Frank Satterthwaite provides a major break-through by which the 20 factors may all be surveyed in as few as 30 test units. Underlying the successful application of Random Balance techniques, we again find under the maldistribution principle, that only two, three, or four of the 20 factors under examination are responsible for a major share of the total variation.

In a chemical company, manufacturing a variety of plastic materials, the largest quality loss was attributable to one particular quality characteristic of one of the products. Multi-vari charts indicated that the greatest variation occurred from hour-to-hour. This could be explained by any of 12 operating parameters including several temperatures, percentages of ingredients, speeds, etc. Six of the 12 were used by the operators to maintain control of the process. The recorded history of the 12 variables measured indicated their normal range of operating conditions. Each of the 12 was programmed into a Random Balance Experiment of 30 test runs with the two levels of each set at the high and low values of their normal operating experience. The results showed that only two of the suspected six, along with one of the unsuspected 6, were critical in their effect on the product. A considerable portion of the variation which still remained was finally explained by instability in the gages in the processing equipment. By maintaining rigid specifications of operating conditions for these three variables, improving instrumentation, setting-up well-defined rules for regular calibration and the elimination of operator judgment in making adjustments, 10% quality losses on the product were reduced to 2%.

Briefly described, the Random Balance Experiment requires the randomization of an equal number of each of the levels of each of the factors among the 30 runs (or other specified number of runs).

Test				Variables				Test
No.	I	II	III	IV	. . . .	XX		Results
1	H	M	H	L	. . . .	C		52
2	L	H	L	L	. . . .	A		36
3	L	L	L	M	. . . .	C		53
4	H	H	L	L	. . . .	B		57
5	L	H	H	L	. . . .	B		24
6	H	M	H	H	. . . .	A		44
7	L	M	L	H	. . . .	B		17
8	L	L	H	M	. . . .	B		37
9	L	L	H	H	. . . .	A		26
10	H	M	L	M	. . . .	A		61
11	L	H	H	L	. . . .	C		37
12	L	L	H	L	. . . .	C		34
13	H	M	H	M	. . . .	B		41
14	H	M	H	M	. . . .	B		21
15	H	L	L	H	. . . .	C		77
.	.	.	.	.	. . . .	.		
.	.	.	.	.	. . . .	.		
.	.	.	.	.	. . . .	.		
29	L	H	L	M	. . . .	C		51
30	H	M	L	H	. . . .	A		37

Figure II

Variable I has 15 High levels and 15 Low levels in randomized order. Variable II has 10 High levels, 10 Medium levels and 10 Low levels in Random order. Variable XX has 10 each of different lots A, B and C of one of the raw materials in Random order.

The results are quickly surveyed in a series of scatter diagrams for each input variable against the measured output characteristic.
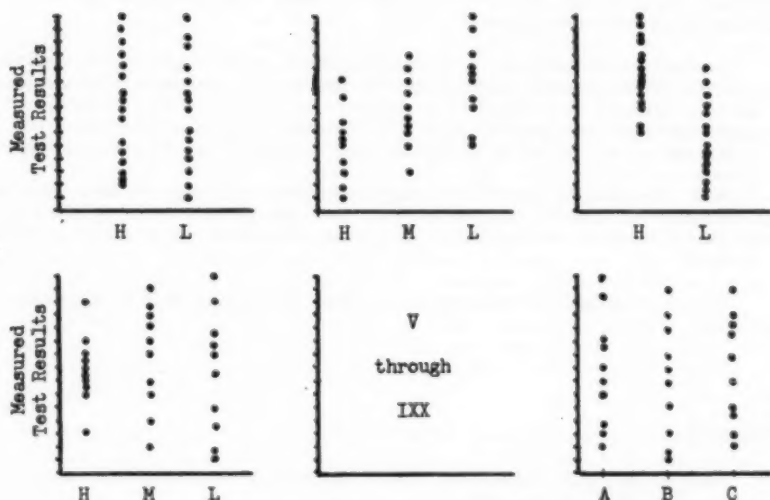
Figure III

It is readily apparent that the difference between the High and the Low levels of Variable I contributes nothing to the total variation. The change from High to Medium to Low levels on Variable II make a definite contribution to the total variation and similarly for Variable III. Thus, less total variation is possible if Variables II and III are restricted within narrower operating limits.

The illustration shows the technique stripped down to its simplest elements. While it is somewhat more involved in design and in analysis, it is a basically simple and attractive approach which quickly gains the confidence of both management and operating level personnel.

One final step may be required in establishing the optimum operating conditions for identified critical variables. The Box technique of evolutionary operation (3) is a sufficiently simple approach which can be understood and appreciated by supervisory personnel. If two critical factors require optimization, small changes in these factors in series of 2 x 2 designed experiments will indicate the direction for optimum conditions until they are finally reached.

When the systematic activity has removed the largest source of quality loss and the second largest and so on down the line, a point must finally be reached at which the reduction of the next largest source of quality loss would no longer justify the expenditure of QC dollars. If the major causes contributing to quality losses, as uncovered, are placed under a reliable system of operating or process control, and if specifications on operating conditions are written or re-written to include the important discoveries, an economic quality level once attained can be maintained at relatively low QC costs.

A complete QC program, as mentioned above, must begin at the earliest conception of the product and follow through to the consumer. The science of Quality Control is sufficiently advanced to supply equally effective programs in every key quality area. QC losses on new products would be at a minimum if effective QC planning were brought in at the ground floor.

When the quality losses are high because of the absence of a complete and effective QC system, Management will not disregard a program based on facts in dollars and cents, but will support one which makes sense as well as dollars.

This is the kind of program which must and can be sold to top management.

REFERENCES:

(1) SEDER, L. A. "Diagnosis with Diagrams" Industrial Quality Control, January 1950.

(2) SATTERTHWAITE, F. E. "Random Balance Experimental Designs", Proceedings Middle Atlantic Conference, A.S.Q.C. Philadelphia, February 1957.

(3) BOX, G.E.P. "Evolutionary Operation". Proceedings of Symposium on Design of Industrial Experiments, Institute of Statistics, Raleigh, N. C., November 1956.

QUALITY CONTROL STANDARDS FOR LABORATORY PRECISION

George T. McCollough, Jr.
Wyandotte Chemicals Corporation

Controlling continuous processing operations and the qualities of flow-type products have created many unique problems for the chemical industry. Special measures of quality are often required that are different from those found in other industries. The filling line for medicinal packaging, or the canning and drumming operations for other chemical products can be evaluated by the standard quality control methods, but the qualities of the bulk material being fed to these end-use packaging operations may not be so readily measured or controlled.

The chemical plant operator has long been aware that frequently chemical laboratory analyses left something to be desired. Sometimes he feels frustrated by the fact that the same raw materials, equipment and procedure today produces a product that rates a different quality result from the control laboratory than did the product of these same factors yesterday. At the same time, the laboratory chemist, equally aware of the same deviations from expected results, can find many reasons why samples can be variables. Phantom causes for differences are created and many misconceptions as to the real value of certain procedures, raw material qualities or other factors affecting the product may arise. In short, the precision and accuracy of chemical analyses have often been misunderstood and usually viewed more favorably than justified by reality.

Chemicals are characterized by their frequent occurrence in large masses, either as a solid, liquid or gas. Both batch and continuous stream production are employed as well as combinations of the two. More often than not a batch of chemical, and particularly if it is a one phase liquid, will be homogeneous within itself, thus making the test precision the only variable pertinent to its evaluation. On the other hand, continuous production streams may be varying constantly, introducing the sample as a potential source of variation. While standard sampling procedures have been developed for many chemical commodities, there is much room for improvement in this important area. The development of appropriate sampling plans for the sampling of heterogeneous chemical streams or masses often requires special adaptation to each individual situation.

The customary measures of quality are augmented in chemical quality control by many more abstract measures such as purity, solubility, pH, density, and the like. Many chemical product specifications contain requirements that special highly empirical tests be passed. Frequently these are attempts to duplicate or estimate in some way the ultimate end-use performance of the product. Such tests as dustiness indexes, foam height, toxicity, thermal stability and corrosivity are examples of these. In these empirical types of testing, the results obtained are usually only meaningful, as they can be compared on a relative basis to the results on other samples or standard materials using exactly the same procedures. Much of the quality evaluation is of the "variable" type with very little of the "go or no go" technique employed.

Sometimes time lag introduced by the test procedure is an important factor in complicating the operator's problem of controlling the process. This is particularly true where the material being checked must continue on to the next processing step without waiting for test results and adjustments as a result of the test. In rare instances, the time to run a test can extend to several days. As a result, more empirical relationships are sought to set up predicting tests which will reduce the risks created from the long test delay times. The precision of such tests as related to the longer test creates new opportunities for variance to enter the picture.

Many chemicals are mixtures containing a relatively high value material diluted with a low value material. For example, in certain grades of sodium carboxymethyl cellulose, an important ingredient in many laundry product formulations, the active agent content may account for only two-thirds to three-fourths of the physical weight of the product, but will account for 96 or 97% of the product's cost. In such cases, sale of the product may be based on a guaranteed minimum content of the higher priced component. The manufacturing cost of the product may be a direct function of a test result. The producer in giving sound value wants to be sure he meets his guaranteed minimum, but he probably will not be paid for the margin of safety he includes. Thus, the quality of the laboratory test evaluating the mixture may make a significant difference in the cost of the product to him. Occasionally the price of the product to the consumer will be calculated from a formula using an analysis result as a factor in computing the price. For example, 50% liquid caustic soda is sold on a basis where the unit price of each shipment is directly proportional to the sodium oxide content of the caustic liquor. Whenever pricing is directly based on the laboratory analysis as in this example, the precision of the analytical procedure becomes extremely important. Smooth vendor-customer relations can be difficult to maintain when the precision and accuracy of the test method are not well understood and agreeable to both parties. The frequently difficult question of "between laboratory" variation can enter the picture at this point, because the producer typically must pass the customer's test as the customer runs it, thus he must be concerned not only with the precision and accuracy of the test in his own laboratory, but also its precision and accuracy in the customer's laboratory.

Historically, the laboratory chemist has armored himself against criticism by running duplicate or triplicate analyses. Typically, these were run from the same sample simultaneously, with identical manipulation and by the same operator. If good agreement was reached, that proved to him that the results were valid and that no errors had been made. The possibility that all the replicates were wrong was, as a rule, overlooked. Plant operating personnel have often resorted to taking new samples for recheck by the laboratory when first results (and their replicates) were not in the expected range. A better value report on the new sample gave everybody some comfort - the laboratory man blamed the sample for the differences and looked no further, and the plant man had the desired answer from the "better sample. The fallacy possible in this situation is obvious to us now that we are more sophisticated as a result of statistical quality control, but it was largely unrecognized until recent years. I might say that in practically every case that I am familiar with, the variance of the chemical test procedure is understated by the results of replicate determinations made under the conditions mentioned above. There is validity in this approach

to reduce the variance of test data, but only when all the random factors affecting the test results are encouraged to be present within the system of replicates rather than be eliminated. This may be accomplished by running the replicate sample by a different person, or in a different laboratory, or on a different day, or a combination of these factors or others suitable to the particular situation.

Overconfidence in the results of laboratory tests on the part of operating personnel can lead to overcontrol of the process. This can have serious consequences for both cost and quality. Many continuous processes have optimum operating conditions, short of which or beyond which efficiencies decline. If tests measuring the relationship of the process to an optimum condition are not precise, they may indicate an adjustment when in reality, none is required. This can lead to a cycling effect which will effectively spread the process performance so it averages at less than the desired optimum. Also, many such adjustments consume labor which could be better used elsewhere. A process that is out of control will create need for more laboratory services and frequently cause off-specification production creating more unnecessary costs.

Summarizing to this point, it is a fact that many problems in chemical processing are created as a result of test imprecision, and that the beginning of wisdom in chemical plant control lies in a clear understanding of the nature of the testing system as it is related to the process objectives that must be met.

Ordinarily, the standard required for a test's precision must be looked on as a derivative of other important factors. First to be considered are the specification limits within which the product must fall, and, secondly, the nature of the real process variability. Both must be in terms of the test being considered.

There are a number of interesting aspects to the process of arriving at specification limits, many of which are quite unscientific and unrelated to the vendor's ability to produce or test, or even to the customer's real need. However, no matter by what process, specifications are set and they are the first reference points for establishing the requirements for test precision.

Examination of the total variance of the system is required. The total variance is that computed from all of the test results on routine production. Abnormal results arising from process upsets due to assignable causes should be eliminated from the estimate of total variance. Resolution of the total variance into its components follows - how much is from real process variation, how much is from sampling, and how much is contributed by the analysis procedure? To be reasonable, the specification must allow at least enough room for an amount of real process variation consistent with the capability of the process. The difference between the variance allowed by the specifications and the amount that will be needed for the real process variation is the maximum that can be tolerated in the analytical system and sampling. This relationship is expressed in the following equations.

Equation 1

$$(\text{Spec. Range}/6)^2 \geqslant \left(\sigma_T^2 = \sigma_P^2 + \sigma_A^2 + \sigma_S^2\right)$$

$(\text{Spec. Range}/6)^2 = $ Total variance allowed by specification range.

$\sigma_T^2 = $ Total variance of system as observed from analytical results on product samples.

$\sigma_P^2 = $ Real process variance which estimates the true variation of product quality about its mean value.

$\sigma_A^2 = $ Variance of analytical procedure which estimates the tendency of the individual test result to vary from the true value of the sample analyzed.

$\sigma_S^2 = $ Variance introduced through sampling which estimates the tendency of the individual sample to vary in quality from the true quality of the material it was taken from.

Real process variation must be stated in the terms provided by the test method. Real process variation is estimated by the difference between the apparent total variance and the amount of variance that can be attributed to sampling and the analytical system as in this equation.

Equation 2

$$\sigma_P^2 = \sigma_T^2 - \left(\sigma_A^2 + \sigma_S^2\right)$$

Since the total variance of the system (σ_T^2) can be estimated from operating data and the determination of variance due to sampling (σ_S^2) can be estimated by an experimental procedure and since both must be stated in terms of the test result, the determination of a practical value for the variance of the test (σ_A^2) becomes the cornerstone upon which the understanding of our process system is based.

Upon inspection of Equation 1, it can be seen that in order to evaluate the total variance of the system as related to the potential variance allowed by the specifications, the sum of real process, sampling, and test variances must be considered. After the values of these have been determined, inspection of the equation will suggest whether any action is necessary, and if so, what direction it may most profitably take.

An estimate of the variance arising from sampling (σ_S^2) can be made by selecting several samples from a typical batch or increment of product and analyzing each several times. Analysis of variance into within and between samples will provide a method for attaining the value required. The design of this experiment must be considered carefully in order that a sufficiently precise estimate of sampling variance is obtained to make the term usable in the total variance equation. Care to replicate the analyses in such a way that full value is given to the within samples variance is important. The inclusion of more than one batch or increment of production may be advisable. Such an experiment will also provide an estimate of the test precision (σ_A^2) which can be used in the over-all equation. In many cases, experience will indicate

300

that the sampling variance will be nearly equal to zero. This is usually more likely in liquid systems, and in such cases, a different approach to estimating test variance (σ_{A}^{2}) can be used.

Considerable thought should be devoted to the relationship of the method of selection of the sample to the character of the process. A flowing fluid, for example, may be varying continuously over a certain range. Instantaneous samples taken from this stream will reflect the complete range of variation at that point in the system. If a surge tank is in the system downstream from the sampling point, mixing will occur and samples from this tank will reflect less variation than was present in the samples of the stream feeding into the tank. Similarly, drip samples or composite samples from the flowing stream will deaden the effect of the variability in the stream. Since the control we are seeking is on the product of the process, the ultimate size of the product unit, which could be anywhere from gallon cans or less to tank cars or larger bulk storage units, should be considered in deciding on the sampling plan. It is easily possible for the sample variability to be less than that among the final product units of the process. To this extent, true process variation will be concealed, and we will obtain a falsely optimistic estimate of the probability of the product unit passing the customer's acceptance inspection.

The most satisfactory situation will be obtained when the samples analyzed are as truly representative of the product and its variability as possible. This will tend to give a σ_{s}^{2} approaching zero and simplify the over-all problem by eliminating a complicating variable.

When the real process variation has been established, the need for improvements in the analytical method can be evaluated. If the real process variation is within the specification limits, but the variance contributed by the analytical method is enough to cause some good product to be rejected or marginally bad product to be passed, method improvement is a possibility. A realistic appraisal may reveal that satisfactory improvement is possible. On the other hand, variance contributed by the analytical procedure can be reduced by obtaining more than one determination on the sample. While care must be taken to avoid the pitfalls of "replicate" testing, variance reduction up to the maximum expected from the formula

$$\sigma_{\bar{A}}^{2} \geqslant \sigma_{A}^{2}/n$$

can be achieved. Because of the fact that time is usually critical, perfect conditions for arriving at independent estimates may not be achieved and $\sigma_{\bar{A}}^{2}$ will be larger than might normally be thought. This is why this equation is written as an inequality rather than in its usual form of an equality.

If sampling variance is significant, the additional tests may be run on additional samples which will tend to bring the average analysis still closer to the true product value. The number of samples and analyses per sample can be selected so as to give the desired probability of error in deciding whether to accept or reject the product.

301

If the real process variation is such that it exceeds the specification limits, obviously some bad product will be produced and the quantity of bad production can be estimated. Here again, inspection of the variance contribution of the various elements will give a clue as to the way to resolve the situation. For instance, a careful examination of the basis for the particular numbers used as specification limits may be advisable. Sometimes, when it becomes obvious that the existing process is incompatible with the product specifications, a change in specification can be arranged satisfactorily. Where this is not possible, and frequently it is not, the need to be exact in separating bad from good product should be evaluated and agreed to by the interested elements in the business. This decision should be based on a knowledge of the test precision, the cost of improving it, the cost of rework of bad product, the possible long run effect of marginal quality deliveries on the long run sales picture and other factors pertinent to the particular problem. The sampling and analysis plan should be so designed as to obtain the agreed upon degree of exactness in appraising the product.

The need for process improvement can be better understood when the real process variation is known as distinguished from the observed total process variation. Usually the cost of reducing process variation increases at some rate above that which a linear relationship would indicate. The probable cost of improved process stability will be more realistically viewed when the contribution of analysis error to the total observed variation is understood.

The standard for control laboratory precision must be considered on a case-by-case basis. The facts of each situation when understood will dictate the precision required for the analytical procedure involved. Precision can be improved basically by one of two methods - getting a better test, or running more results per sample. Which course is the better to follow will be principally a function of the cost of the test and the technology available for improvement.

As a continuing technique for evaluating and controlling test precision or variance we have used the following simple program. Samples run by the analyst are commonly set aside as "retainer" samples for rechecking should some future development indicate a need for re-analysis. The laboratory leader feeds back to the analyst at intervals an unknown sample which is in reality a portion of the "retainer" sample he has previously tested. His recheck analysis is made without his knowing which of the samples he previously tested he is running, though he may realize he is checking himself. A graph of the range of the original and recheck result is maintained in the laboratory. This is kept as a modified range chart with plus and minus values being plotted. (See example Figure 1.) Plus and minus values for the range are obtained by subtracting the recheck result from the original and keeping the sign of the answer. R-bar for computing control limits is obtained by averaging the individual ranges as if they were all plus values, as they ordinarily would be considered. The usual upper control limit for range, for sample size equal to two, is charted both as a plus and as a minus value. From R-bar, the variance of the analytical method is estimated.

Figure 1

Figure 1
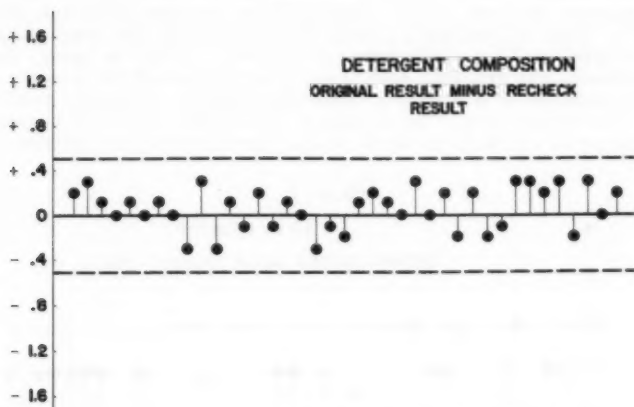
In addition to providing the needed estimate of the analytical variance, this type of test rechecking and charting have provided interesting trouble-shooting values, and their maintenance is primarily justified as a laboratory test quality control device. Figure 2 shows a chart prepared from the original data taken through this recheck procedure on the method shown in Figure 1. Note that initially the limits for range were very wide compared with the later time period of the chart. The reason for the improvement in precision in this case was the result of investigation and modification of the test method. This improvement was mandatory because the strength of a detergent product was being controlled by it. The variance of the analytical method was much larger than that allowed in Equation 1 when allowance was made for real process variation inherent in the system.

Figure 2

303
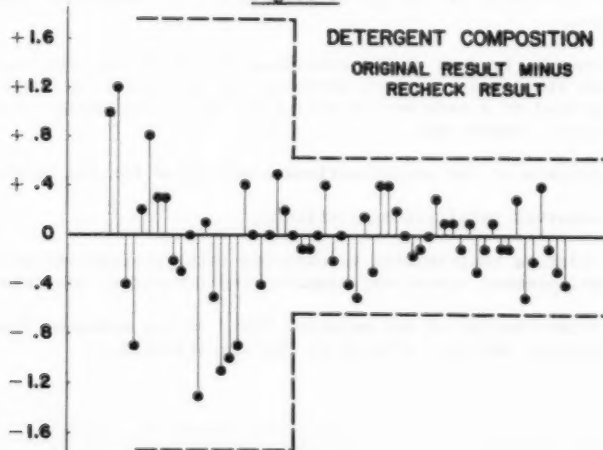
Another example of a problem brought to light by the recheck pro-
cedure is illustrated in Figure 3. Here, an obvious bias to the plus
side indicates that the A.P.I. gravity value obtained from the sample
is usually lower on rechecking than was obtained on the fresh sample.
Such a consistent bias could come from product change on aging, but in
this instance, the original sample was usually tested for gravity while
hot and corrected to a standard temperature while the second test was
run at room temperature. The standard correction value for temperature
was found to be slightly in error which accounted for the difference.

Figure 3

API GRAVITY

ORIGINAL RESULT MINUS RECHECK RESULT
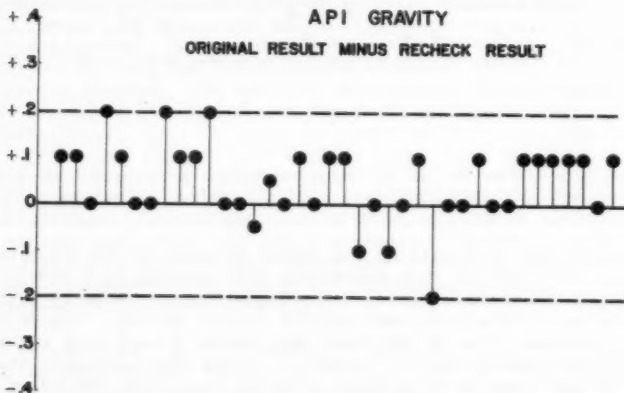
In most cases, an operator rechecks his own original work.
Plotting all operators' results on the same chart helps to point up the
quality of the individual's performance. Statistically, this intro-
duces the possibility that between operator differences exist that are
not reflected by R-bar. Where this is suspected, cross-operator re-
checks will rectify the statistical estimate of variance, but will
limit the usefulness of the chart as a control of individual operator
performance.

In summary, because of the considerable possibility for variance
in chemical laboratory testing, establishing the standards for the pre-
cision required of a test method will require the evaluation of several
other factors. These are:

1. An analysis of the specifications established for the product.

2. The observed total process variance.

3. The relative contribution to observed total process variance made
    by real process variation, sampling and laboratory analysis.

4. The determination of the relative "fit" of the variance obtained
    in practice and that allowed by the specifications.

5. The exactitude with which acceptance or rejection decisions are to be made on product approaching specification margins.

6. The scale of the penalties both short and long range that will accrue as a result of making incorrect quality decisions.

STOPPING RULES FOR CONTINUOUS PRODUCTION

R. B. Murphy
Bell Telephone Laboratories, Inc.

In continuous inspection poor quality over an extended period may lead to more inspection effort than is desired. The producer, consumer, or both may feel that economic and quality interests could be well served at such times by resort to a special action aimed at improving the process. There has been considerable interest recently in providing criteria based on inspection data to determine when such action should be taken. Such criteria have been called stopping rules because in many cases production is stopped when the special action is called for.

The present discussion concerns four such stopping rules as applied to Dodge's original CSP-1 plan, in which inspected units must be classified as either "defective" or "nondefective". The basic plan starts with inspection of each unit made in succession until a specified number, i, of consecutive units are found nondefective. A sequence of units so inspected is called a screening sequence. After the initial screening sequence is ended, the inspector samples a fraction f of the units presented to him. He continues to sample until he finds a defective unit. At this point screening is resumed, and the procedure is repeated indefinitely. The inspector rejects (or sets aside for correction) any inspected unit found to be defective and accepts all others.

Of interest are the practical aspects of applying the four stopping rules as well as operating characteristics of these rules which may serve as a guide to selecting a suitable version for application.

TABLE I

STOPPING RULES FOR CSP-1 PLANS

Rule ($n^* - i$):   Stop as soon as a defective unit is found in any one
screening sequence after the sequence has exceeded
$n^* - i$ units.

Rule (r):   Stop as soon as a specified number r of defective units
are found in any one screening sequence.

Rule (N,R):   Stop as soon as a specified number R of defective units
are found in any block of a specified number N of in-
spected units. (Blocks do not overlap.)

Rule (n^*):   Stop as soon as a specified number n^* of units have been
inspected in any one screening sequence without ending
it.

TABLE II

DEFINITIONS AND NOTATION

p = The probability that a defective unit is produced = Process average.

$q = 1 - p$.

i = The number of nondefective units produced and inspected in a row before sampling may be used.

f = The average fraction of output inspected when sampling is used.

F = The total fraction of units inspected.

PNQL = Producer's Nominal Quality Level, a value of p chosen with the AOQL and F to fix f and i.

T_n = The probability that no run of i nondefective units has occurred by the time n units have been produced. (NOTE: For brevity the expression for T_n is omitted.)

$k = (1-T_{n*})/(1-q^i T_{n*} - T_{n*+i})$ = Correction factor for characteristics of Rule $(n*)$. (See Table IV.)

$D_r = (1 - q^i)^r$ = The probability that no run of i nondefective units has occurred by the time r defective units have been produced.

$B_n(x) = \sum\limits_{m=x}^{n} \binom{n}{m} p^m q^{n-m}$ = The probability that no more than n units have been produced by the time x defective units have been produced.

Stopping cycle = The elapsed production period from one time a given stopping rule is invoked until the next time it is invoked.

Counting cycle = The elapsed production period from one time defectives begin to be counted until the next time they begin to be counted in accordance with a given stopping rule. (Applies only to Rules (r) and (N,R).)

m_P = The (variable) number of units produced in a stopping cycle.

m_C = The (variable) number of units inspected in a counting cycle.

$E(m_P)$ = The average of m_P.

$E(m_C)$ = The average of m_C.

TABLE III

ILLUSTRATIVE CSP-1 PLANS AND STOPPING RULES

Specified				
AOQL	1%	1%	5%	5%
PNQL	.005	.005	.025	.025
F at PNQL	.10	.25	.10	.25
$E(m_p)$ at PNQL	10^4	10^4	10^4	10^4

CSP-1 Plan				
i	150	70	30	15
f	.05	.20	.05	.20

Actual				
AOQL	1%	1%	4.8%	4.6%
PNQL	.005	.005	.025	.025
F at PNQL	.1004	.2620	.1011	.2677

Rule $(n^* - i)$				
n^*	251	127	79	36
$n^* - i$	101	57	49	21
$E(m_p)$	10,920	11,520	10,000	6,877

Rule (r)				
r	2	2	4	3
$E(m_p)$	10,900	11,290	9,701	6,707
$E(m_C)$	305.8	259.2	78.64	56.63

Rule (N,R)				
N	400	175	97	44
R	3	3	5	4
$E(m_p)$	11,990	11,260	9,737	6,821
$E(m_C)$	356.7	172.1	94.99	43.78

TABLE IV

EXPRESSIONS FOR F, E(mp), AND E(mc)

	F	E(mp)	E(mc)
Basic CSP-1 Plan	$f/(f + (1-f)q^i)$	-	-
Rule $(n^* - 1)$	"	$(1-T_{n^*})/pq^i FT_{n^*}$	-
Rule (r)	"	$(1-D_r)/pq^i FD_r$	$F \cdot D_r \cdot E(m_P)$
Rule (N,R)	"	$E(m_C)/FB_N(R)$	$N - B_R(R) - B_{R+1}(R) - \ldots - B_{N-1}(R)$
Rule (n^*)	$f/(f+(1-f)q^i k)$	$(1-T_{n^*})/pq^i FT_{n^*}k$	-

QUEUEING THEORY AND SOME OF ITS INDUSTRIAL USES

Roger I. Wilkinson
Bell Telephone Laboratories, Inc.

Delays in giving service occur whenever in any process or system, the momentary demand exceeds the available capacity of the serving personnel or equipment. Familiar examples are people waiting in the checkout line at a supermarket, at the railway ticket window, and in a doctor's office. Automobiles waiting to cross arterial highways and toll bridges, airplanes seeking to land at busy airports, and ships awaiting dock space to unload are examples of costly and critical delay situations in the transport field. In industry there are such diverse waiting line problems as the flow of materials and machinery to assembly and test positions, the extent to which inoperative machines should be required to wait for repairs, and the economic backlog of typing work in a central-ized stenographic department. In all of these the waiting "customers" comprise a queue. Queueing theory is the study of the formation of such customer waiting lines and the characteristics of the delays they experience.

The earliest mathematical studies of delays were made by A. K. Erlang of the Copenhagen Telephone Company; he published his theories over a period of years beginning in 1909 [1]. Other well known contribu-tors to the field include E. C. Molina [2], T. C. Fry, C. D. Crommelin [3], F. Pollaczek [4], C. Palm and J. Riordan, all of whom were strongly concerned with the flow of traffic through telephone switching systems. Since World War II the techniques of queueing theory have become an important tool of the workers in Industrial Engineering and the popular Operations Research field. Hundreds of articles have now been written on queueing theory and its applications.

A short derivation of the Erlang formulas is given in the Appendix. The assumptions made in this analysis probably correspond more closely than any others to the preponderant number of applications so far made in industry. References 5, 6 and 7 give bibliographies of papers containing the development of theories and formulas for a very wide variety of practical and theoretical situations. T. L. Saaty [8] has recently made a résumé of useful queueing formulas; and P. M. Morse [9] has provided a readable and comprehensive analytical approach to general queueing problems.

Elements of the Queueing Process

In the upper part of Fig. 1 are shown pictorially the principal elements required to establish a queueing situation. At the left are a number N of Sources whose Demands arise at successive instants of time. These comprise the Input Process*. If, when the demands reach the Servers - assumed to be c in number - all servers are found busy, the demands will normally wait, along with any others already there, until their service moment arrives. When a server becomes available, a choice is made of some one of the waiters to be accommodated. The particular manner of making this choice is called the Queue Discipline. There will

- -

*The terminology used here is considerably based on that of D. G. Kendall.

then be a Service Time during which the demand is satisfied, following
which the demand is dismissed. The point in time character of the dis-
missed demands is called the Output Process. The latter is of particular
interest when the output is all or part of the input to a second queueing
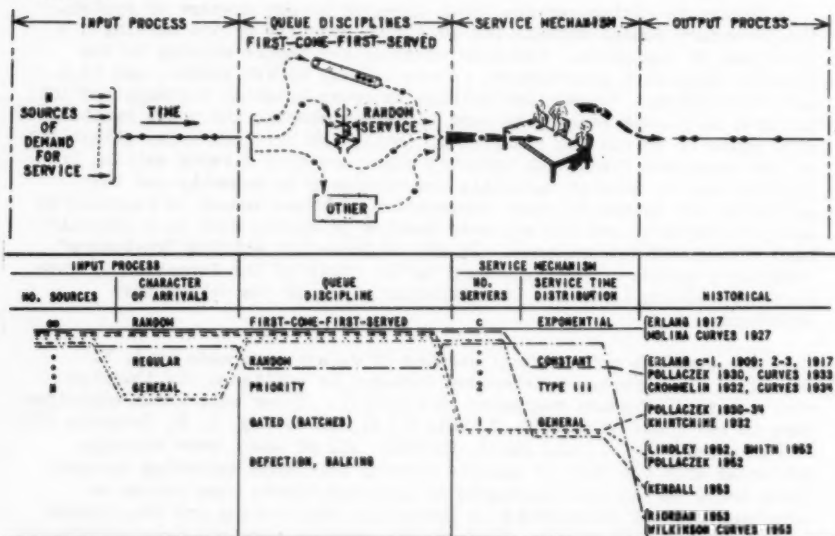system.

Fig. 1 - Display of Queue Elements, and Common Assumptions

In the lower part of Fig. 1 is a table displaying the most commonly
encountered choices for the several elements of the queueing process.
Under Input Process the selection of N very large, or infinite, simpli-
fies the analysis since then the arrival rate of demands is unaffected by
the number of sources already engaged in waiting or being served, i.e.,
the demand rate is constant.

The Character of Arrivals may be highly diverse. The two simplest
cases are those where the demand instants are at random (their inter-
arrival times have an exponential distribution), and where they arrive at
regular intervals (constant interarrival time). The most general assump-
tion would of course leave one free to select any desired distribution of
interarrival times.

In delay systems involving people, first-come-first-served is
commonly accepted as the fairest rule by which successive selections are
made from a line of waiting customers. However, attainment of this
Queue Discipline may require substantial expense. In this event a random
choice for the next-to-be-served from among those waiting may occasionally be
acceptable. Certain classes of demands (customers) may merit Priority
over others who have arrived earlier (e.g., emergency telephone calls).
One can imagine circumstances in which last-come-first-served might be
observed (cf. a new baby in the house!). Customers served in Batches are
not uncommon (buses, taxis). Defections from the queue may occur if the

314

delay becomes too great, or refusal to enter the queue (Balking) may be the result of seeing too long a waiting line.

The distribution of Service Times is critical in determining the distribution of delays experienced by the demands. It is very common in real life (e.g., repair of machinery, retail customer purchasing, telephone conversations) to find Exponentially-distributed Service Times,

$$\varphi(t)dt = \frac{1}{h} e^{-t/h} dt \tag{1}$$

Again, in many machine operations a Constant Service Time occurs. More specialized service time distributions are sometimes described by Pearson's Type III curve, while, of course, a general frequency function will describe service times of any character [1], [5].

In the last column of the table of Fig. 1 are given several historical references of practical importance to appliers of queueing theory; at the end of a line associating assumptions made in successive columns, is the author's name and the date at which he published his findings. Further details are found in the list of references.

Criteria of Delay Service

Depending on the circumstances, interest in the results of the queueing process may take different forms.

The Probability of Delay, $P(>0)$, is the simplest index of delay service. In instances where the penalty is great for being delayed at all, selection of a small acceptable value of $P(>0)$ may suffice to engineer or administer a system. Thus a delay in supplying a necessary part in an assembly line might cause a process to stop with attendant confusion and large costs of starting up again.

Probability of a Delay Greater than t, $P(>t)$, will often be required especially if some secondary event tends to occur as delays exceed certain critical amounts. Waiting lines of customers, who begin to complain, or even defect, if the delays seem unreasonable, are of this sort. Again, airplanes with limited fuel which encounter landing delays have a strong interest in the probability of delays beyond certain limits.

Average Delay, $\bar{t}$, is required to be known when the aggregate of many delays, rather than the magnitude of any individual delay, is important. Thus a factory manager might want to estimate the total lost machine time (and hence production), due to machine stoppage when repair men or facilities are not always immediately available (due to being engaged with other stopped machines).

Probability of Queue Length Exceeding q, $Q(>q)$, will be desired when special provision for handling the queue is required. Thus the number of seats in a waiting room, or buffer storage units in a computer, will need to be sufficient to strongly insure that no more than a very few demands are turned away.

Average Length of Queue, $\bar{q}$, is needed to estimate the aggregate lost time of waiting demands. If employees waiting for a within-the-business

service, average too great a delay time, it may be economical to hire
more servers and reduce the average length of queue.

Mathematical Formulas

Let a = average simultaneous demands being served (if none defect)

= number of demands per average service time h

$$= \frac{(No. \ bids \ per \ hour) \ (Avg \ service \ time \ per \ bid \ in \ secs.)}{3600} \quad (2)$$

(Note that the unit of load "a" is appropriately called the erlang.)
The "occupancy" level, or load per server, is then defined as $\alpha = a/c$.

For random (Poisson) input, no defections, and exponential service
times (and very closely so for constant service times), the probability
of a particular customer being delayed, i.e., meeting a delay greater
than zero - as derived in the Appendix, eq. 13 - is given by

$$P(>0) = \frac{(a^c e^{-a}/c!) \ [c/(c-a)]}{1 - \sum_{x=c}^{\infty} a^x e^{-a}/x! + (a^c e^{-a}/c!) \ [c/(c-a)]} \quad (3)$$

Values of $P(>0)$ are graphed on Fig. 2 (taken from [10]), whose
abscissa is load a; the numbers on the curves correspond to the numbers
of servers c. For example, a load of a = 3 erlangs offered to five
servers results in a probability of delay of 0.236, that is 23.6% of
customer demands can expect to find all servers busy and thereby have to
wait for service.

Delays Greater Than t, Exponential Service Times

The probability $P(>t)$ of a customer demand being delayed longer than
t is the product of two probabilities: (Probability of being delayed
at all) $\times$ (Probability, when delayed, of being delayed longer than t).
This may be written

$$P(>t) = P(>0) \ F(u) \quad (4)$$

where u = ct/h. Here the presence of u clearly introduces the delay time
t, the number of servers c and average service time h into the solution;
one would expect the result to depend on each of these factors. $P(>0)$
values are read from Fig. 2. For first-come-first-served situations,
values of $F_Q(u) = e^{-(c-a)t/h}$ - derived as equation (20) in the Appendix -
are read from Fig. 3. For cases where waiting customers are served in
random order, there is no corresponding simple expression for $F_R(u)$;
approximate values, however, may be read from Fig. 4.

Example 1: Delays at Airplane Ticket Counter. If in the average busy
hour 48 customers purchase airplane tickets at a two-position sales
counter, what proportion will have to wait more than 5 minutes?
Suppose service times are roughly exponential and average 2 minutes.
Assume (a) first-come-first-served, and (b) customers served at
random.

316
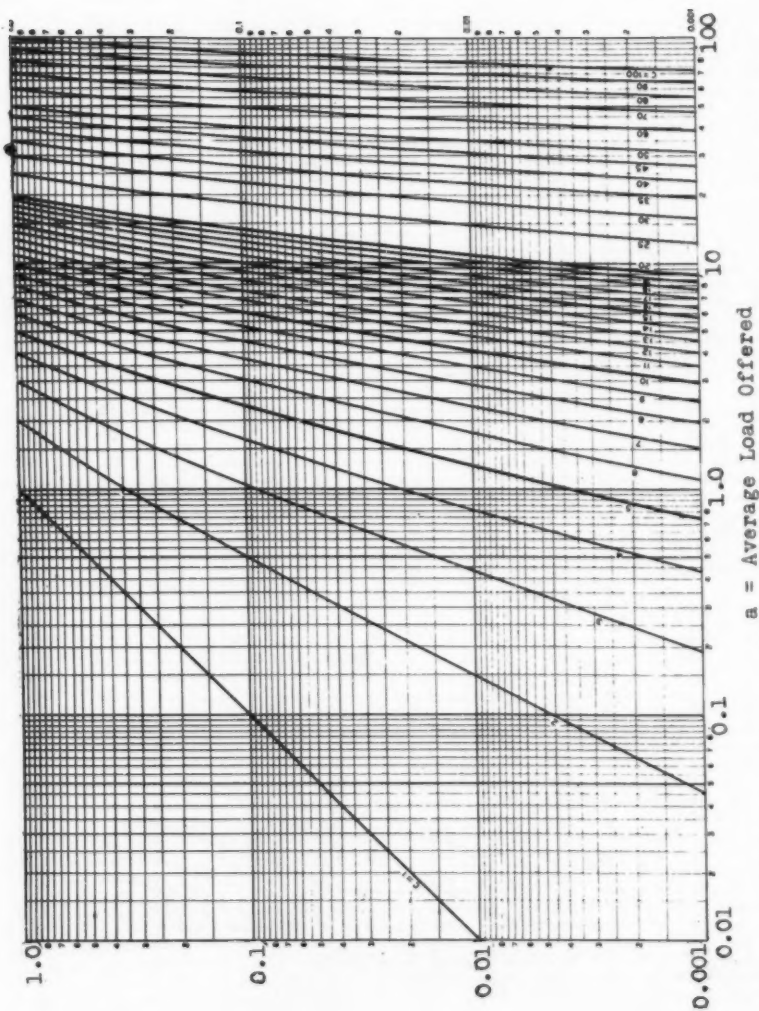
Fig. 2 - Probability of Delay in terms of Number of Servers c, and Offered Load a.

a = Average Load Offered
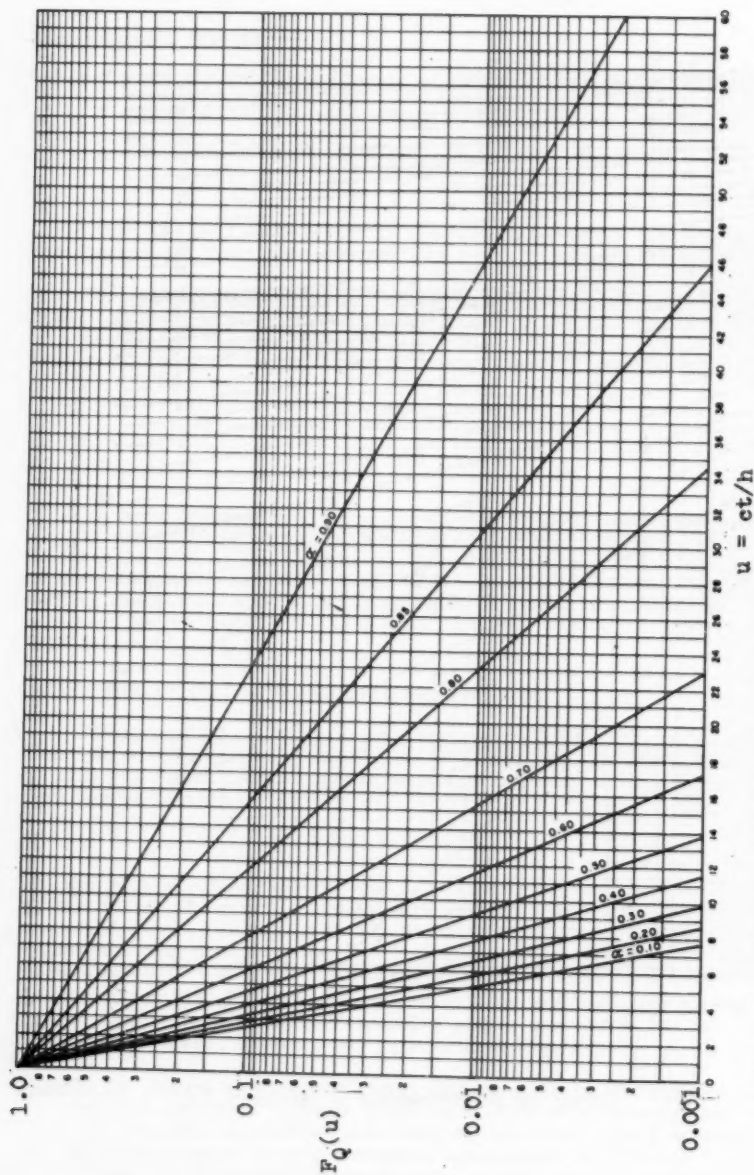
P(>0) = Probability of Delay

317

Fig. 3 - Probability of <u>Delayed Exponential</u> Demands being Delayed Longer than t, when served in <u>Queued</u> (First-Come-First-Served) Order.
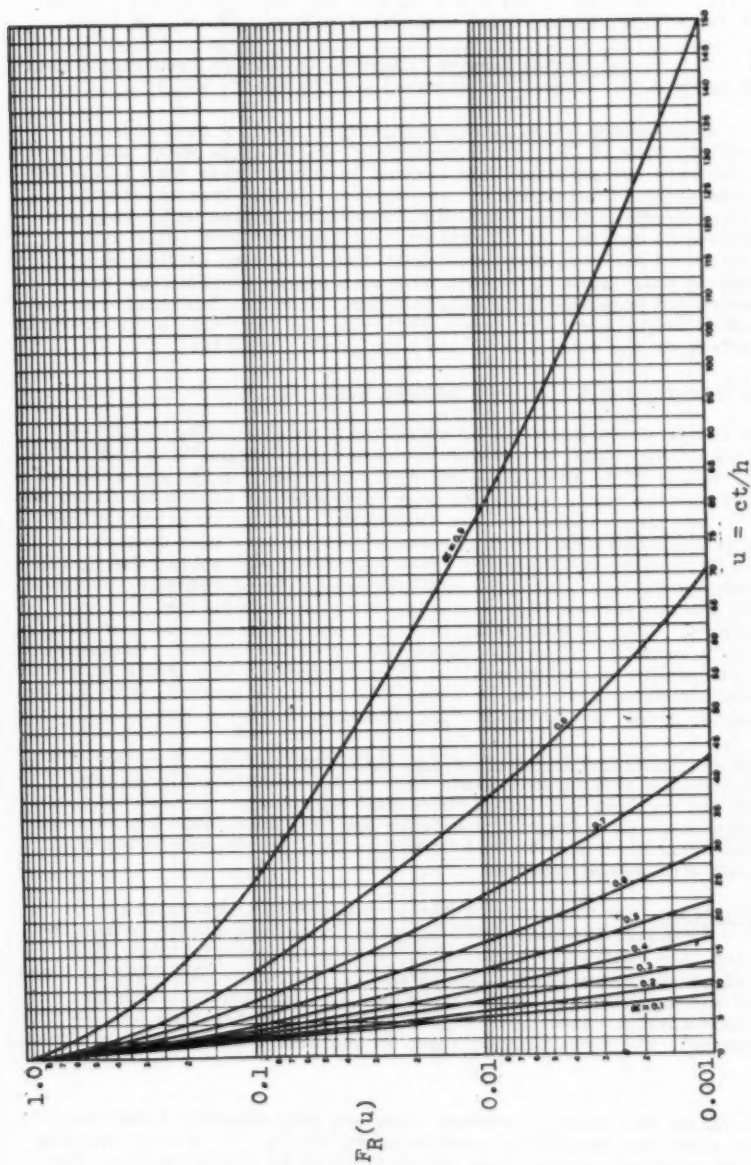
318

Fig. 4 - Probability of Delayed Exponential Demands being Delayed Longer than t, when Served in Random Order.

319

Solution: (a) Queued Service. a = nh/3600 = (48)(120 secs.)/3600 secs = 1.6 erlangs; α = 1.6/2 = 0.8. From Fig. 2, $P(>0)$ = 0.71. From Fig. 3, using u = ct/h = (2)(5 min.)/2 min. = 5, read $F_Q(u)$ = 0.37, whence $P_Q(>5$ minutes) = (0.71)(0.37) = 0.26.

(b) Random Service. Same as (a) except from Fig. 4 read $F_R(u)$ = 0.27. Then $P_R(>5$ minutes) = (0.71)(0.27) = 0.19.

As suggested above, the probability of finding both ticket sellers busy (0.71) is independent of the order in which waiting customers are served. This however is not the case when delays of some specific magnitude are considered. In the present example a third fewer customers must wait longer than 5 minutes if they are served at random! At 10 minutes, the same per cent would wait; beyond 10 minutes, more would wait than if queued service were maintained. It will be seen that with random service more very short and more very long delays occur. Thus the length of delay at which the criterion of acceptability is set, may have a considerable effect on the selection of the best queue discipline.

Delays Greater than t, Constant Service Times

Although the probability of suffering a delay at all, $P(>0)$, is relatively insensitive to the service time distribution,* the probabilities of delays $P(>t)$ when t>0, are highly dependent on the character of the service time variations. The only case which is widely tabulated to date, in addition to that for the exponential service times considered in the previous section, is for constant service time with demands handled first-come-first-served. Even here the mathematical theory is surprisingly difficult and does not lend itself to constructing a simple pair of charts (such as Figs 2 and 3), covering a wide field of loads and servers. (For constant service times and random order of service, a reasonably computable formula remains yet to be developed.) Crommelin[3] first published a set of charts for selected numbers of servers to c = 20, and occupancies to 0.8. Two of these, for c = 1, 2 servers, with higher occupancies added, are shown in Figs. 5 and 6.

Example 2: Telephone Engineering for Dial Tone Service. A relay-type controller serving 2000 lines in a certain automatic telephone system requires 0.5 second to connect dial tone. If each line originates an average of two calls per busy hour, what per cent of the calls will have to wait more than 1, 2 and 3 seconds for dial tone with queued service?

Solution: The occupancy of the single server (controller) is α = (2000)(2)(0.5 sec.)/3600 secs. = 0.556. Since with no congestion, the controller requires 0.5 second to give dial tone, the delays of interest are 0.5, 1.5 and 2.5 seconds. Expressed in terms of average service time, they become 1, 3 and 5 units of delay, respectively. Reading on Fig. 5 (constant service time for c = 1 server), the probabilities shown in the last column below are found.

- -
*$P(>0)$ values for constant service times are very slightly lower than the values given for exponential service times on Fig. 2. A chart showing the true constant holding time values is given by Pollaczek [4]. For most engineering purposes the difference is negligible.
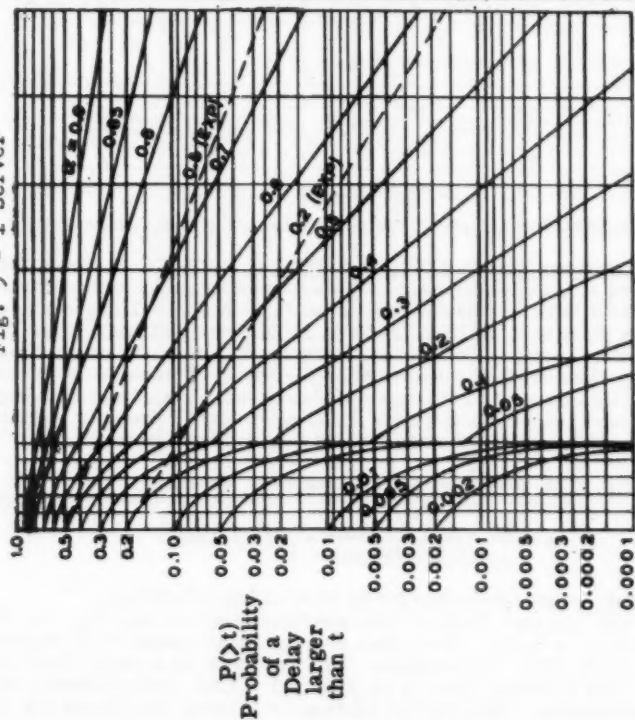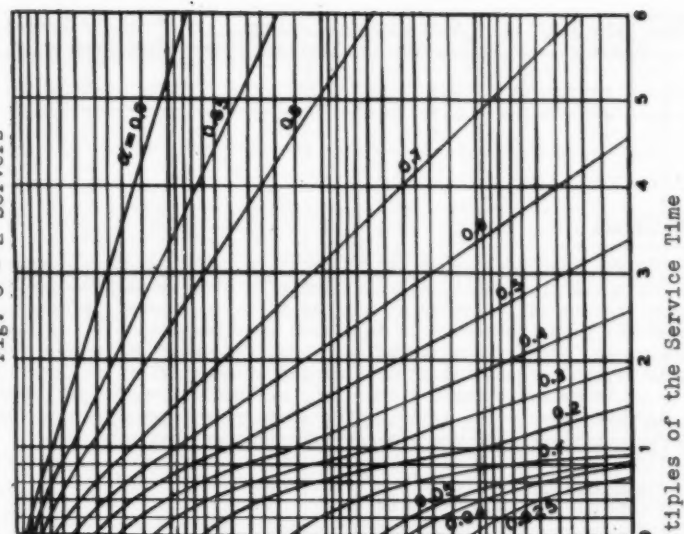
320

Fig. 5 – 1 Server

Fig. 6 – 2 Servers

Probability of __Constant Service Time__ Demands being Delayed Longer than t, when
Serviced in __Queued__ (First-Come-First-Served) Order.

321

Dial Tone Time	Corresponding Delays		Prob. of Exceeding
Seconds	In Seconds	In Service Times	Dial Tone of Col. 1
1	0.5	1	0.22
2	1.5	3	0.026
3	2.5	5	0.003

Average Delays

In many situations, principal interest is in the average length of delay. As noted previously, this is likely to occur when the cumulation of delays represents lost time or expense to an individual or company, as distinguished from the annoyance felt when some particularly long delay is encountered. It may readily be seen that the order of service from the queue is inconsequential in determining average delay, as long as there is no tendency to favor either long or short service time demands in the service order. The distribution of service times, however, has a considerable influence on the average delay, as we shall now see.

For exponential service times the average delay - derived as equation (18) in the Appendix - is

$$\bar{t}_E = P(>0) \, h/(c-a), \qquad (5)$$

while for constant service times it is [3]

$$\bar{t}_C = h \sum_{w=1}^{\infty} e^{-aw} \left[\sum_{v=wc}^{\infty} \frac{(aw)^v}{v!} - \frac{c}{a} \sum_{v=wc+1}^{\infty} \frac{(aw)^v}{v!} \right] \qquad (6)$$

These two expressions are represented on Figs. 7 and 8, respectively.

Example 3: Optimum Size for Oil Truck Filling Station. The owner of a fleet of oil delivery trucks wishes to estimate how much delivery time his trucks are losing due to waiting at his filling station, and how much this would be reduced if he installed additional filling units. Suppose he now has 28 trucks and four filling pumps, and it requires 20 minutes to fill a truck. If each truck needs three fillings on the average per 8-hour day, what is the average truck time lost waiting at the filling station? How much waiting time would be saved if he installed two or three more pumps?

Solution: The "load" on the filling pump system is

$$a = \frac{28 \times 3 \times 20 \text{ mins.}}{8 \times 60 \text{ mins.}} = 3.50 \text{ erlangs.}$$

With four pumps the occupancy is $a = 3.50/4 = 0.875$.
Enter the constant holding time average delay curves, Fig. 8, with $c = 4$ and $a = 0.875$. Read that average delay equals 0.75 holding time = 0.75 (20) = 15 minutes. For the fleet as a whole then the daily lost trucking time is 15 x 28 x 3 = 1260 truck minutes, or 21 truck-hours. Thus due to inadequate filling facilities the time of nearly three trucks, on the average, is being wasted.
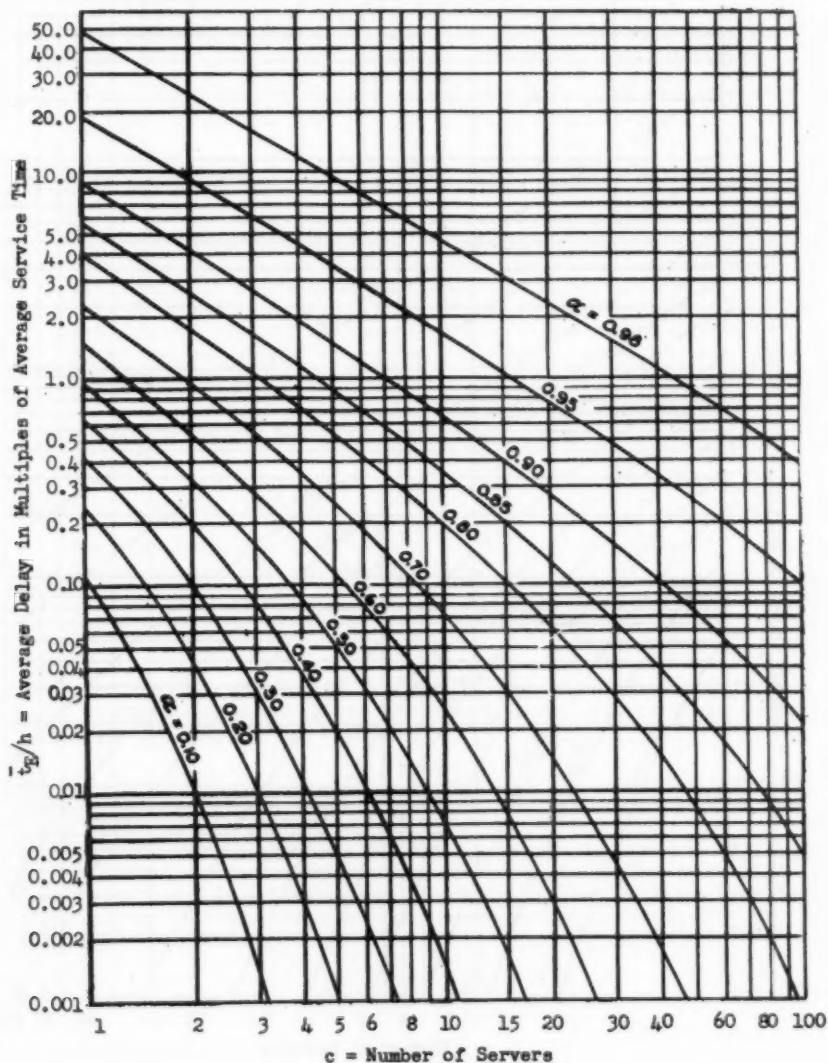
FIG. 7 – AVERAGE DELAY ON ALL DEMANDS, WITH <u>EXPONENTIAL</u>
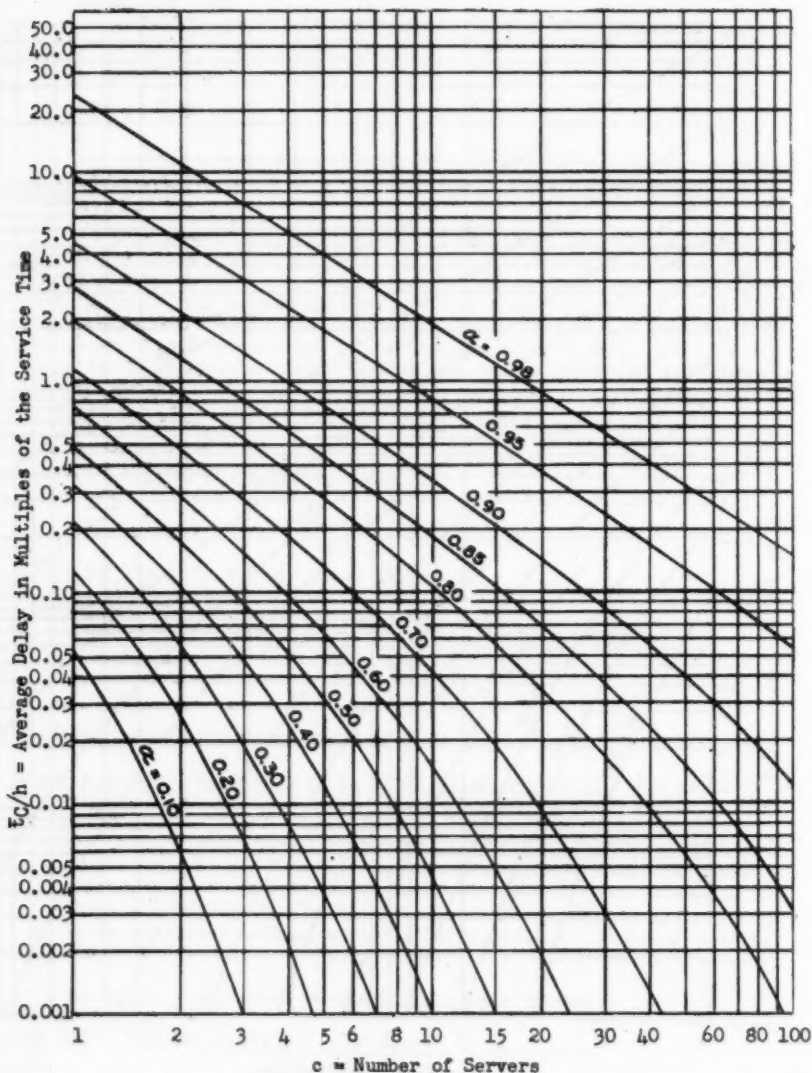SERVICE TIMES.

323

FIG. 8 - AVERAGE DELAY ON ALL DEMANDS, WITH <u>CONSTANT</u>
SERVICE TIMES.

324

The following table shows the improvement in lost truck time if r additional filling pumps are installed:

Added Pumps r	Total Pumps c + r	Pump Occupancy	Avg. Delay $\bar{t}_C$	Daily Lost Time (truck hrs)
0	4	0.875	15.0 mins.	21 hrs.
1	5	0.700	2.7 mins.	3 hrs. 47 mins.
2	6	0.583	0.78 mins.	1 hr. 5 mins.
3	7	0.500	0.27 mins.	23 mins.

Two small corrections might well be applied to this example. The actual delays will not be quite as great as those predicted here since infinite source theory has been used to solve a system with only 28 sources. A correction should also probably be made to allow for slightly improved delivery schedules as the number of pumps is increased, and with correspondingly increased demands by each truck for refilling at the pumps.

It seems likely that the oil distributor would conclude that he should install two, and possibly three more pumps (business might increase).

Length of Queue:

For exponential service times the probability of a queue of length q or longer, is shown in the Appendix (equation 14) to be

$$Q(\geq q) = \left(\frac{a}{c}\right)^q P(>0) \qquad (7)$$

while the average queue length is $\bar{q} = P(>0) \frac{a}{c-a}$. Moreover the average delay on demands delayed, $\bar{t}$, and the average queue length during periods of all servers busy, $\bar{q}$, are related by

$$\bar{\bar{t}}_E = \frac{\bar{\bar{q}} + 1}{c/h} \qquad (8)$$

in which $\bar{\bar{t}}_E = \bar{t}_E / P(>0)$ and $\bar{\bar{q}} = \bar{q}/P(>0)$.

Equation (8) is satisfactory for exponential service times only. For constant service times, average queue length may be estimated from the approximate relationship

$$\bar{\bar{t}}_C \doteq \frac{\bar{\bar{q}}_C}{c/h} + \frac{h}{c + 1} \qquad (9)$$

or its equivalent

$$\bar{q}_C \doteq c \left(\frac{\bar{t}_C}{h} - \frac{P_C(>0)}{c + 1}\right) \qquad (10)$$

where $\bar{t}_C/h$ is read from Fig. 8, and $P_C(>0)$ may be approximated by the exponential service time values, $P(>0)$, of Fig. 2.

325

Example 4: Factory Production and Test Facility Provision. In a certain factory, complex machine assemblies come off several production lines at irregular intervals, averaging 42 per hour, and are then sent to one of eight test positions for acceptance or rejection. On the average it requires 10 minutes to test a machine although due to their intricacy and the adjustment possibly required, this is a highly variable time, found to be roughly exponential.

(a) If machines are stored on dollies awaiting test, how many dollies should be provided so no oftener than once in 500 cases will there by no place to park a completed machine?

(b) All production is required to be passed through test at the end of the day. After 5 P.M. when production stops, the cost of testing increases by $15 per waiting machine. How much could be spent per day toward the purchase and operation of an additional test position?

Solution: Assume random arrival of demands, and exponential service times; then $n = 42$, $c = 8$, $h = 10$ minutes, so that $a = nh/60 = 7.0$ erlangs; and $\alpha = a/c = 0.875$. From Fig. 2, $P(>0) = 0.64$; by equation (5), $t_E = h/(c-a) = 10$ minutes.

(a) Set $0.002 = Q(\geq q) = \left(\frac{a}{c}\right)^q P(>0) = (0.875)^q(0.64)$. Solving for q gives $q = 43.2$; hence 44 dollies should be provided.

(b) At 5 P.M. (as at any other time during the day) the expected number waiting in the queue is

$$\bar{q} = P(>0)\frac{a}{c-a} = 0.64 \frac{7}{8-7} = 4.48$$

If another test position were installed $(c = 9)$, then

$$\bar{q}' = 0.38 \frac{7}{9-7} = 1.33$$

The average daily savings in overtime testing would be $(4.48 - 1.33) 15 = 47.25 which could be spent toward an additional test position and operator. Note also that the number of dollies required in (a) would be greatly reduced (to 21).

Summary

The elements of simple queueing problems have been displayed and some formulas given by which probabilities of delays and expected delays may be calculated. Working curves are shown by which many common queueing problems can be either exactly or approximately solved. Examples are given which illustrate the use of the curves in practical situations.

326

Selected References on Applied Queueing Theory

1. E. Brockmeyer, H. L. Halstrom and Arne Jensen: The Life and Works of A. K. Erlang, Copenhagen Telephone Company, 1948.

2. E. C. Molina: "Application of the Theory of Probability to Telephone Trunking Problems," Bell System Tech. Jl., Vol. 6, p. 461, (1927).

3. C. D. Crommelin: "Delay Probability Formulae," Post Office E. E. Jl., Vol. 26, p. 266, (1934).

4. F. Pollaczek: "Kurven für die Gesprächsverluste im vollkommenen Leitungsbündel," Elecktrische Nachrichten Technik, Vol. 10, p. 396, (1933).

5. D. G. Kendall: "Some Problems in the Theory of Queues," Jl. Roy.Stat. Soc., Vol. 13, No. 2, p. 151, (1951).

6. Vera Riley: "Bibliography of Queueing Theory," Operations Research for Management, Vol. II, Johns Hopkins Press, 1956, p. 541.

7. Alison Doig: "A Bibliography on the Theory of Queues," Biometrika, Vol. 44, p. 490 (December 1957).

8. T. L. Saaty: "Résumé of Useful Formulas in Queueing Theory," Jl. of Opns. Res., Vol. 5, p. 161 (April 1957).

9. P. M. Morse: Queues, Inventories and Maintenance, New York, John Wiley and Sons, 1958.

10. R. I. Wilkinson: "Working Curves for Delayed Exponential Calls Served at Random," Bell System Tech. Jl., Vol. 32, p. 360, (1953).

Derivation of Formulas for
Exponential Holding Time Delays

Let c = number of servers
   a = rate of demand arrivals per average service time
   t = time of delay
   h = average of exponentially distributed service times
   $f(x)$ = probability of x simultaneous demands at any random instant

Assume:

1. Number of sources is very large compared with servers so arrival rate is independent of the number of sources being served or waiting.

2. Demands which are not served immediately, wait and are served in order of their arrival.

If a<c, the system will attain stable equilibrium, and the equations relating the state probabilities are:

For $0 \leq x \leq c$, $(a+x)\ f(x) = af(x-1) + (x+1)\ f(x+1)$, or $f_1(x) = \dfrac{a^x}{x!}\ f(0)$

For $x \geq c$, $(a+c)\ f(x) = af(x-1) + cf(x+1)$, or $f_2(x) = \dfrac{a^x}{c^{x-c}c!}\ f(0)$

$$\left. \right\} \quad (11)$$

Since $\qquad \displaystyle\sum_{x=0}^{c-1} f_1(x) + \sum_{x=c}^{\infty} f_2(x) = 1,*$

$$f(0) = \frac{1}{\displaystyle\sum_{x=0}^{c-1} \frac{a^x}{x!} + \sum_{x=c}^{\infty} \frac{a^x}{c^{x-c}c!}} = \frac{e^{-a}}{\displaystyle\sum_{x=0}^{c-1} \frac{a^x e^{-a}}{x!} + \frac{a^c e^{-a}}{c!}\ \frac{c}{c-a}} \qquad (12)$$

- -

*If at some queue length b, customers refuse to join the waiting line (i.e., balk), the upper limit of x will be c+b instead of infinity. As a result $\dfrac{c}{c-a}$, wherever it appears in equations (12) and (13) will be modified by the factor $\left[1-(\tfrac{a}{c})^{b+1}\right]$. The probability of balking will, of course, be simply $f_2(c+b)$.

Probability of Delay

The probability that any randomly chosen demand will find all servers busy and be delayed (more than zero time) is

$$P(>0) = \sum_{x=c}^{\infty} f_2(x) = \sum_{x=c}^{\infty} \frac{a^x}{c^{x-c} c!} f(0)$$

$$= \frac{\dfrac{a^c e^{-a}}{c!} \dfrac{c}{c-a}}{\displaystyle\sum_{x=0}^{c-1} \frac{a^x e^{-a}}{x!} + \frac{a^c e^{-a}}{c!} \frac{c}{c-a}} \tag{13}$$

Note that all terms except $\frac{c}{c-a}$ are either Poisson individual or summation terms, so that evaluation of $P(>0)$ is easily accomplished. Values of $P(>0)$ are read from Fig. 2.*

Queue Length, $Q(\geq q)$ and $\overline{\overline{q}}$:

The probability of finding a queue of length q, or longer, is

$$Q(\geq q) = \sum_{x=q+c}^{\infty} f_2(x) = \sum_{x=q+c}^{\infty} \frac{a^x}{c^{x-c} c!} f(0)$$

$$= \left(\frac{a}{c}\right)^q \sum_{x=c}^{\infty} f_2(x) = \left(\frac{a}{c}\right)^q P(>0) \tag{14}$$

The average queue length is

$$\overline{q} = \sum_{x=c}^{\infty} (x-c) f_2(x) = f(0) \frac{c^c}{c!} \sum_{x=c}^{\infty} (x-c)\left(\frac{a}{c}\right)^x$$

$$= f(0) \frac{c^c}{c!} \left(\frac{a}{c}\right)^c \frac{\frac{a}{c}}{\left(1-\frac{a}{c}\right)^2} = P(>0)\frac{a}{c-a} \tag{15}$$

The average queue length when all c servers are busy is, then,

$$\overline{\overline{q}} = \frac{a}{c-a} \tag{16}$$

- -
*Tables of $P(>0)$ under the expression $E_{2,n}(A)$ are given in Table III of MOE'S PRINCIPLE, A. Jensen, Copenhagen Telephone Company, 1950 for ranges of offered loads A = .1(.1)10, 10 (1)50, 52(4)100, and E = 0.0001 to approximately 0.6000.

Average Delay, t_E:

When there is a delay, the delay equals the wait for (number in queue + 1) demands to be served. During periods of all-servers-busy, calls are served at a rate of c per average service time h. Then the average delay for calls delayed is

$$\bar{t}_E = \frac{\bar{q} + 1}{c/h} = \frac{\frac{a}{c-a} + 1}{c/h} = \frac{h}{c-a} \tag{17}$$

and the average delay on all calls is

$$\bar{t}_E = P(>0)\ \bar{t}_E = P(>0)\ \frac{h}{c-a} \tag{18}$$

Distribution of Delays, $P(>t)$:

The probability of a random demand finding m calls in the queue is $f_2(x = m+c)$. Then the demand must wait for m+1 calls to drop out of the system. When all-servers-are-busy this occurs at a rate of c per average service time h; or ct/h per time t. The probability of the demand having to wait longer than time t, equals the probability that fewer than m+1 will drop out in time t, or

$$p_m(>t) = \sum_{r=0}^{m} \frac{(ct/h)^r e^{-ct/h}}{r!}$$

The total probability of a delay $>t$ results from summing for all m:

$$P(>t) = \sum_{m=0}^{\infty} f_2(m+c)\ p_m(>t) = \sum_{m=0}^{\infty} f(0)\ \frac{a^{c+m}}{c^m c!} \sum_{r=0}^{m} \frac{(ct/h)^r\ e^{-ct/h}}{r!}$$

$$= f(0)\frac{a^c}{c!}\ \frac{c}{c-a}\ e^{-(c-a)t/h} = P(>0)\ e^{-(c-a)t/h} \tag{19}$$

And the probability of a delayed call being delayed $>t$ is

$$F_Q(u) = e^{-(c-a)t/h} \tag{20}$$

where $u = ct/h$. Values of $F_Q(u)$ are read from Fig. 3.

DEVELOPMENT OF QUALITY ATTITUDE AND PRIDE OF WORKMANSHIP

Harry E. Sagen
Abbott Laboratories

Quality attitude, quality mindedness, and quality consciousness are used synonymously, but quality attitude implies something more than just being aware of quality.

Even before the early industrial revolution, craft guilds rose and fell on the quality of workmanship. Pride of workmanship with quality attitude manifested through the artisans product was the keynote of the day. Throughout the years, the successful craftsman must have constantly strived for quality by the very reason for his continued existence.

In those days, the Acceptance Quality Level was truly determined by the customer, while the Average Outgoing Level was uppermost in the mind of the skilled craftsman who passed on through his apprentices, the skills and knowledge he had attained. Through all of this training, there was continuous emphasis upon quality with the development and maintenance of quality attitude and pride of workmanship.

Improving and maintaining quality in the process of reducing cost and price in the field of fair competition plays an important role that must be a part of the daily work habit of all employees, from the top management of the organization down through the ranks to, and including, those who do the most menial tasks which are classed as non-skilled.

You have undoubtedly attended a large fair or livestock exposition and watched the judging of animals which is a form of inspection, and also have watched the exhibition of showmanship by both 4H Club Groups and seasoned adult exhibitors. The exhibitor as well as the judges can only measure quality through attributes that can be observed by sight and feel. The exhibitors are mindful of this, and use all their ethical methods possible to emphasize these attributes of quality and use their cunning showmanship to accentuate the attributes associated with quality. In all the preparation of feeding, grooming and showing, the exhibitor has an ultimate goal, the grand championship with its accompanying joy in the pride of accomplishment. Have you ever seen anything more gratifying than the smile of pride on the face of a winner? My only reason for bringing in the foregoing example is, because all of you have at some time or other experienced the same joy in pride of workmanship which culminated because of a quality attitude held during the period in which you accomplished your goal.

To bring about the development of "Quality Attitude and Pride of Workmanship" calls for the use of all the principles, knowledge and techniques in the teaching and training of employees, and all the principles embodied in good human relations in the handling of employees.

To develop a quality attitude and a pride of good workmanship calls for an intelligent program, put on by personnel having an understanding of human behavior. Alexander Pope, in his essay "Essay on Criticism" gives us an excellent guide in his statement, "Men should be taught as if you taught them not, and things unknown be proposed as things

forgot." This can be paraphrased in many ways, but in a very humble way tells us to forget about the personal pronoun "I," and to encourage doing on the part of others by sharing and even giving up some of our coveted recognition. The true recognition will eventually find the deserving person who has willfully dropped the use of the personal pronoun "I."

The techniques and types of programs best suited to develop a quality attitude and pride in workmanship will depend in part upon the size and character of the organization and industry, the size and character of the working sub-group, and the range in characteristics of the individuals.

Our greatest aim must be to restore to the individuals, the pride of workmanship they had before the days of mass production. Through the years of our industrial history, we have developed the idea that we can influence persons in large groups if we have the right program and literature. This is in part true, but if we lose sight of the individual and personal basis, we lose the effect of the binding kingpin, the individual who can unintentionally make or break the program to be developed. To really influence large groups in the development of quality attitude, one must deal with the individual on a personal basis. Any program that is to be used to influence workers should be geared to take in the individual.

In the Training Within Industry Service, developed during World War II, the Job Relations Training emphasized the fact that supervisors get results through people. The foundations for good job relations were listed as; (1) let each worker know how he is getting along, (2) give credit when due, (3) tell people in advance what changes will affect them, and (4) make best use of each persons ability. To build best upon this foundation, people must be treated as individuals.

Now then, what can we do if we are to treat the workers as individuals in a program of developing quality attitude and pride of workmanship? Suppose we assume that top management does not have to be sold on the value of a good quality control program in which job relations play such an important part. If this is true, the efforts can be expended in developing an attitude among the line staff and workers, whether they be in plant or in office operations.

The personnel you have on hand would not be exchanged except, insofar as to make the best use of each persons ability. However, the personnel responsible for the selection and employment of people should through their interviewing and pre-employment testing in aptitudes and attitudes, evaluate the prospective employee in terms of job and quality attitude. Many employees are classified as accident risks. In a like manner, workers tend to classify each other on the basis of attitude and workmanship risks. If this information can be obtained from periods of previous employment of the new worker, it will be of material help to the supervisors during the training and orientation periods.

The training and orientation program should emphasize development of a quality attitude and pride of workmanship with the same emphasis that safety, reduction in operating cost, employee health, and other points are now stressed as factors in the over-all picture of plant operations and handling of people.

332

Quality attitude and pride of workmanship are developed through accomplishment, primarily through proper handling of people. I am reminded of the youngster in the third grade in school who had just completed some written homework for the following days class. It was done in a very slip-shod fashion, with erasures and crossing out of words, which the parent suggested warranted re-copying to make a paper worthy of being handed to the teacher. The youngster, however, questioned the value of this effort and made the statement, "Why should he care how the paper looked as long as the teacher accepted that kind of work." Did not the statement of the youngster perhaps reflect the real attitude and lack of pride in workmanship on the part of the teacher? Contrast this example with another class from the same school a few years later under the influence of teachers who developed a feeling of pride in accomplishment among the pupils. This group gained an enviable reputation for neatness, completing of assignments, and accuracy through the setting of high standards of quality, and which when attained gave the pupils a pride of workmanship. Have we not seen examples similar to these in industry and everyday life? There is no difference in developing attitudes in industry from that in the classroom. The incentives are not always the same, but the principles of handling the individuals are the same. There is an inherent dignity among individuals, in which man's instinct to be a part of a group in which he can take pride in accomplishment is a factor to be recognized.

The worker takes pride in belonging to an organization where the day-to-day actions of top management stimulate confidence and respect for the sincerity of their actions. This is true in regards to the day-to-day actions which reflect quality attitude, as well as those reflecting other attributes which are characteristic of a strong, live, organization.

The individual on the assembly line takes pride in his work and company when he knows the product he is putting out represents the best efforts and knowledge of research, development, engineering, manufacturing, control, purchasing, advertising, and sales. Effective employee relations are maintained with the minimum effort when the individual feels the pride of belonging; of belonging to an organization which has earned the right of loyalty to the company, devotion to the supervisors, and a pride of workmanship, because of the policies and actions which have gained the respect and confidence of the individuals.

Perhaps we in the Pharmaceutical Industry are fortunate, and have an enviable position in the effort of developing quality attitude and pride of workmanship among the workers. This is because the management and workers alike, are mindful of the moral and ethical obligation to society through their products, in addition to operating under the legal standards set by the Food, Drug and Cosmetic Act, as well as specific requirements through N.I.H., A.R.S., U.S.P., and N.F. The Pharmaceutical Industry is alert to its responsibilities and obligations to the public and medical profession. Through its various professional and industrial organizations, it is constantly keeping quality, its improvement and control, ever uppermost in mind during their deliberations. Through committees such as the U.S.P., Contact Committee, various committees in the A.D.M.A., A. Ph. M.A., etc., it has taken the initiative. Standards of quality set through regulations have become minimum requirements, just the same as in a good safety program the safety code becomes the minimum standard. The improvement above minimum

standards is up to the individual managements.

A U.S.P. Sub-committee (4-13-49) in stating the general principles of Quality Control in the Pharmaceutical Industry said, "Pharmaceutical Quality Control is the operation of a system of checks and safeguards set up in a plant to assure integrity, strength, quality and purity of the finished products. The purpose is to assure that all merchandise shall meet professional requirements, all legal standards, and in addition, such standards of refinement and elegance as the management of a firm may adopt. It is the sum and substance of all effort which differentiates the manufacture of drugs and medicines from most other manufacturing projects, and makes of it an altruistic profession on a par with the practice of medicine."

In any discussion of Quality Attitude and Quality Control, it must be remembered that the turning out of a quality product is not governed entirely by gauges, analytical balances, control charts, etc. A quality product is not the result of the manufacturing process alone. In this I am sure you will agree. You will agree that a fine pharmaceutical product could not be turned out, even with control chart methods, under unsanitary working conditions. An intricate instrument could not be properly assembled in a poorly lighted factory. Quality attitude must be evident in every phase of a business, from the production of the raw materials right down to the time the final product is used by the consumer.

We are all familiar with the statement that quality cannot be inspected into a product. It must be built into the product.

To build a quality product we are dependent upon management, men, materials, methods and machines. In dealing with the five "M's" the individual remains as the keystone to be fitted into the structure which is asked to support and produce the quality product to satisfy the boss, the final customer. It is no wonder that today the efforts expended to have an educated and enlightened individual as a worker are increasing.

Much has been said and written about the American Economic System and the American Way of Life. Bringing the facts about this great heritage to the supervisory personnel, in order that they may better know the underlying principles of cause and effect, will do much to stimulate pride in our American Way of Living and pride in our accomplishments as a people of the North American continent.

There is an alertness on the part of the management as to the importance of all phases of operations of any company, but it is not necessarily true that this is shared by all the line staff. Likewise, the knowledge of the relation and part played by materials produced or services rendered by the individual is not shared by the workers in the production of a product.

The lack of knowledge of the importance of the individuals's work is often responsible for the various attitudes some not conducive to quality products.

It is utterly impossible to have every person know the functions of all the other people in the various departments. Providing an

opportunity for those who are interested in hearing about how the work of other departments affect their operations, is a means of stimulating work interest. Through this work interest a feeling of pride of belonging is nurtured, from which pride of workmanship and quality attitude can be developed.

In the development or production of a pharmaceutical product, many departments are involved. The ultimate consumer acceptance of the product, based upon quality as manifested in its various attributes, must be their goal. This can be attained when Research, Development, Engineering, Purchasing, Production and Control think and act with a Quality Attitude. A breakdown of this attitude by any one group will jeopardize the contributions of the other groups in the production of a quality product. Dosteoyevski once said, "If it were desired to reduce a man to nothing it would be required only to give his work the character of uselessness." Conferences which are democratic; not dictatorial, autocratic, or sectarian, bring about the discussions which lead to quality attitude and a quality product.

In building a quality attitude and pride of workmanship, there is one group that is often neglected in recognition of its contribution to quality products. This group is none other than the maintenance work force. Basically they have the responsibility of maintaining equipment in condition for the maximum production of high quality material. Foremen are prone to give a time limit for a change-over or repair, which in many cases results in poor quality maintenance, not as a choice on the part of the maintenance worker and set-up man. The maintenance worker is one who should be tactfully handled to develop a feeling of pride of workmanship based upon accomplishment. One must remember that he deals with people under the stress of quantity production. One minute he is dealing with an eager beaver who expects to become plant superintendent over night, and the next minute he deals with an old employee who never believes in change. There are few others in the organization who are so closely associated with cause and effect in production of quality goods. Look well to his suggestions of help for improvement.

The effect on quality and quantity production is very evident when machines fail or fall into dis-repair. Quality Control can help prevent the maintenance man from falling into dis-repair with respect to his quality attitude and pride of workmanship, through its recording of machine performance for future evaluation.

Just as machines perform better when the rated work load is not exceeded, quality production is likewise a greater probability when the human element is not overloaded because of faulty planning and scheduling. Pressure exerted because of extreme necessity for quantity production will usually be followed with increased cost of production due to increase of rejections. Lack of normal work load is also to be avoided. Busy hands and busy minds are essentials in development of quality attitude and a pride in workmanship.

Perhaps we should stop at this point and ask ourselves, what is an attitude? Finlay, Sartain and Tate in their book entitled "Human Behavior in Industry" state, "an attitude is a tendency to act with respect to a certain value." It is an inclination we have for a thing, person, or idea.

335

We cannot always judge an attitude by how a person acts because sometimes the true attitude is suppressed. Attitudes are under the surface and latent; possessing potential energies that can be directed to positive value under the guidance of a basic understanding of human nature. Some workers whom we consider non-conformists in quality attitudes today, are that way simply because they have not as yet attached a positive value to this concept. It is up to us in management and supervision to help the individual acquire the positive value through contact with our experiences. An attitude is not hereditary. It is acquired through contacts.

If attitudes are acquired, can they be changed? Attitudes can be changed through the change in value system, because as the values change our attitudes change.

Many positive values as quality attitude, have been determined by intellectual reasoning. The positive value attached to a quality product is something that the worker can grasp as a logical conclusion through directed reasoning. He knows there is a calculated risk in skating on thin rubber-ice on the pond. He can reason that proper lights are essential for night driving.

Intellectual values are the easiest to change. Emotional attitudes are more difficult to change because with emotions it is how we feel about a thing rather than how we think about it. The workers best intellectual processes may tell him he is wrong, but he may still stick to his feeling about a thing.

Development of quality attitude and pride in quality should be acquired through the intellectual process. Pride in quality of a product is a value to management as well as to workers on the line. The executives like to feel they are associated with the best product in their field. If this feeling and attitude toward quality is contagious, it is reflected in the work habits of the line staffs down through the ranks. The quality minded executive will not compromise quality because he has an attitude toward the value of the product. To the short-sighted, this may appear as a reduction in profit, but to those who see beyond the first horizon, there is the picture of harvest yet to be reaped.

Those who work with people should know that this thing we call attitude is capable of changing. It becomes our duty to lay out the course. People react emotionally. Attitude like morale is difficult to create within a group, but sometimes so easy to lose because we forget the group is made up of individuals differing from one another in experiences which make up the values that can spell success or failure in quality production.

In the development of morale, "the feeling of belonging," the individual feels the welfare of the group is more important than his own. If we can team up morale with pride of workmanship, a combination would be formed that would be hard to lick. To maintain this combination day after day, calls for an intelligent supervisory staff backed by a management cognizant of developing "the feeling of belonging."

Each one of you can perhaps recall instances in which a supposedly weaker team made up of younger unknown players, who had never made

336

success, captured the championship at the end of the season from the better known stars. Constant hustle and a will to win were the characteristics attached to the winners. This is an example of morale, a feeling of belonging; belonging to a group with a definite attitude, and a desire to act toward certain values. If, in our workers, this value is quality attitude and pride of workmanship, the morale will be high because the desire to win in this economic competition through constant hustle and alertness with intellectual reasoning will be an ever present incentive.

You perhaps are disappointed, that in this discussion no definite program has been advanced for obtaining quality attitude. Perhaps this is more intentional than accidental, because the program devised should be geared to the particular organization and its individuals.

In the beginning of my discussion, attention was focused upon the individual. An intelligent and enlightened individual with enthusiasm for his work makes the best raw material for development of quality attitude.

At Abbott Laboratories, employees at different levels have availed themselves of the opportunity to learn about Quality Control and Statistical Methodology, as an adjunct to Research, Production and Control.

All the effort to develop quality attitude has not been expended in formal class training of a few. The success of Safety Training in industry has been due in large part to the constant publicity through posters, meetings, suggestion systems, special drives, etc. Quality attitude and pride of workmanship can be built soundly through the same methods. Recently, in our plant, following a very splendid two-day rally and demonstration of safety methods and equipment, the Production Quality Control Group placed "Push" and "Pull" signs upon various doors, propagandizing the "push" and "pull" for better quality and reduced cost. This idea borrowed from another company, met with much favorable comment because it had the effect of focusing attention upon quality and reduced cost to the same extent to which safety has been accustomed.

In the beginning of this paper, mention was made of the importance of working through individuals. People are to be handled differently from machines, which after reading the instruction for operation can be handled very much alike. With human beings one must study, adjust, and set the examples by doing as you would have them do to develop the enthusiasm for your program.

A program with an objective, clearly in mind, is the start. No use running until you know where you are going. This perhaps can be illustrated by telling the story about the boys attempting to make straight paths across a field of freshly fallen snow. One after another the boys stepped carefully along, placing one foot in front of the other, but with no great success. Another boy came along and with great confidence said he could make a better and straighter path than had been done. He set out at a fast, unhesitating pace, and when his path was completed, it was found to be straighter than that of the others. Pressed for the secret of his success, he replied, "Do you see that oak tree over on the hill? I just kept my eyes right on that and walked toward it without paying attention to the snow or anything around me."

This boy knew in what direction he was going. We, too, should know where we want to go.

Let me close this paper by again quoting from the Essay on Criticism, "But where's the man who counsel can bestow, Still pleased to teach, and yet not proud to know."

ELECTROPLATING APPLICATIONS OF QUALITY CONTROL

W. E. Jones
Management Controls
Des Plaines, Illinois

Electroplating is a word that has not been discussed in the Quality Control literature. A search through Industrial Quality Control revealed no articles on this subject. A search through 1956 and 1957 Abstracts of Statistical Quality Control was also fruitless. W. C. Geissman and R. A. Carlson have presented material on "Current Distribution in Barrel Plating. A Statistical Study", Proceedings, American Electroplater's Society, 1952. Material presented here has been taken in part from work done with Geissman & Carlson at the National Lock Company, Rockford, Ill. Other than zinc barrel plating case studies are from unpublished material from various companies.

The first six charts are limited to the particular field of zinc barrel plating. Additional charts touch on cadmium barrel plating. Comments will include case histories in the fields of galvanizing continuous strands of wire, copper plating, chrome plating, and nickel plating involving all types of plating equipment. Contents of this paper are designed to stimulate further use of a new approach to the economic control of electrochemical plating processes. This paper is not designed to present a complete story on any single process. Much of the material presented on zinc barrel plating is restricted to only one barrel size at one revolution speed in one standard zinc solution.

Barrel plating has often been regarded as a process that is very uniform since all of the parts are tumbled which randomizes the exposure of every part to solutions and current densities. It is almost a universally accepted illusion that barrel plating produces absolutely uniform plating thickness on every part. Statistical Quality Control methods are relatively new to electroplaters. Statistical analysis is the most effective method for investigation of the many variables affecting electroplating quality.

Zinc barrel plating results were obtained from a fifteen barrel unit, consisting of three tanks with capacity of 3 barrels each. Barrels were horizontal, hexagonal, 30 inches long, 14 inches across flats, with flexible dangler contacts, 21mm perforations in Lucite panels, all made by same manufacturer, and all rotated at six revolutions per minute. Standard procedures were used for cleaning and pickling parts. Power supply was a 4,000 ampere, 12 volt generator for each tank. Plating solution formula was: zinc cyanide 10 oz/gal, total sodium cyanide 16.5 oz/gal, and caustic soda 14.5 oz/gal. Sodium sulfide was added to reduce metallic impurities. An organic brightner, RH-309, supplied by the Du Pont Company was also added to maintain brightness. Solution temperature varied from 70 to 90 degrees F. Plating was followed by two rinses but no bright dip. Zinc barrel plated parts were a large variety of wood, machine, and sheet metal screws. Plate thickness was determined by drop test method, Hull-Strausser ammonium nitrate, using an automatic dropping unit made by the Kocour Company. This instrument delivers the standard 100 drops per minute. Identical areas on different pieces were always measured. On screws and bolts, heads were always measured.

Five groups of samples were generally taken from each barrel, spaced evenly along the axis of the barrel. This stratified method of sampling

permitted an analysis from end to center to end of barrel. There was no significant differences in the averages or the variances between groups of samples by location along the axis of the barrel. This permitted substantial reduction in sample sizes taken from each barrel. At the start of the study, 60 samples were taken from each barrel. This was later reduced to 25 samples from each barrel.

Chart 1 shows the barrel average plate thickness plotted against the total variation within each barrel (6 standard deviation value for variance within barrel). This proves that the variation within each barrel increases significantly as the average plating thickness increases. The line of best fit defines the relationship between these two variables. These results are only for the previously mentioned formula solution. Discussion of these results suggest further investigations on the effect of reducing the zinc cyanide progressively down from 10 oz/gal to as low as 3 oz/gal, keeping other elements of the solution about constant. With caustic soda and total sodium cyanide at about $2\frac{1}{2}$ to 3 times the zinc metal, it has been claimed that it is possible to run barrel plating so uniform that it is almost impossible to detect variation on the parts within a barrel. I have not found any published facts showing effect of such departures from "standard" solution formulations. It is also suggested that barrel revolution speeds be changed. Better uniformity has also been claimed at 3 revolutions per minute in 20 minutes instead of 6 revolutions in 30 minutes.

Chart 2 shows no correlation between surface area in barrel load and total variation within each barrel. Chart 3 shows very little correlation between plating time and total variation within each barrel. Chart 4 shows slightly more correlation between ratio of time/area and total variation within each barrel.

Chart 5 is a simplified presentation showing minimum plate thickness assured in relation to average plate thickness. If a minimum specification of .1 mil must be met, it is necessary to plate an average of .22 mils in order to guarantee that 99.87% of all the loads will have less than .1% under specification. To guarantee a minimum of .2 mils requires an average of .37 mils which is almost 90% more zinc than would be required if all parts could be plated right down at the minimum. Most zinc barrel plating processes will range from 90% up to as high as 200% thickness above the minimum of .2 mils required. Production personnel were able to use this chart to reduce the average amount of excess zinc. At the same time, they obtained more uniformly plated parts for shipment. An Operations Research study is suggested at this point for consideration of all the economic factors involved in making the decisions necessary to reach optimum profit from the zinc barrel plating process. Any change in the process that reduces the within barrel variation immediately permits a reduction in the average amount of zinc plating needed to guarantee a minimum. These savings must be weighed against the costs of changing the process.

Chart 6 shows the correlation between ratio of time/area to barrel average zinc plate thickness. Chart 7 shows the correlation between amperage and average cadmium plate thickness. These two charts are typical of many simple comparisons that can be made to evaluate relationships existing on actual production equipment. Advanced multiple correlation and design of experiments are considerably more efficient techniques than the linear correlation studies graphically shown in this paper. These
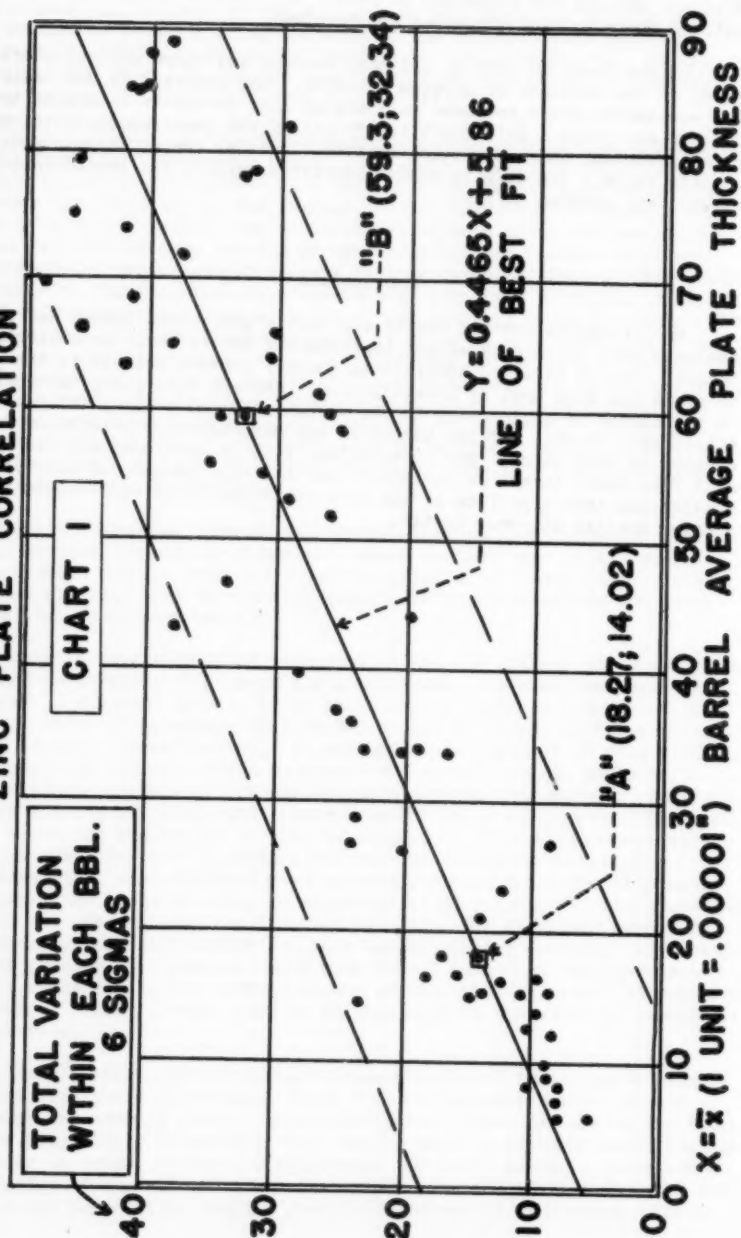
advanced methods are essential to reduce the cost of analytical investigations where so many variables are involved.

Charts 9 and 10 show an ordinary average and range control chart study of the strength of cadmium solution. One analysis is made weekly. Two successive weeks are used as a sample size of two to determine the average and range. This chart is typical of the improvements obtained on all solution characteristics charted: namely; copper oz/gal; sodium cyanide oz/gal; tin oz/gal; sodium hydroxide oz/gal; Ph; sodium carbonate oz/gal; and cadmium oz/gal.

Conclusions

Use of simple control charts and easily understood linear correlation analysis will produce excellent improvements due to their psychological impact. Use of advanced statistical quality control methods is essential to avoid the high cost of investigation. Average zinc plate thickness is correlated to an increase in the variation of zinc plate thickness within each barrel, for the solutions and manufacturing conditions described in this paper. New formulations will be developed and new ideas will find their impact on equipment that has not changed for many years. Statistical investigations in the electroplating field will result in "better quality at lower costs".

$Y = 6\sigma$ BARREL AVERAGE THICKNESS V.S. VARIATION WITHIN BBL.
ZINC PLATE CORRELATION

CHART I

TOTAL VARIATION WITHIN EACH BBL. 6 SIGMAS

"B" (59.3; 32.34)

$Y = 0.4465X + 5.86$
LINE OF BEST FIT

"A" (18.27; 14.02)

$X = \bar{x}$ (1 UNIT = .00001") BARREL AVERAGE PLATE THICKNESS

SURFACE AREA
v.s.
VARIATION WITHIN BARREL

CHART 2

343

PLATING TIME
v.s.
VARIATION WITHIN BARREL

PLATING TIME
1 UNIT = 1 MIN.

TOTAL VARIATION
WITHIN EACH BARREL
1 UNIT=.00001"

CHART 3

344

TIME/AREA
v.s.
VARIATION WITHIN BARREL

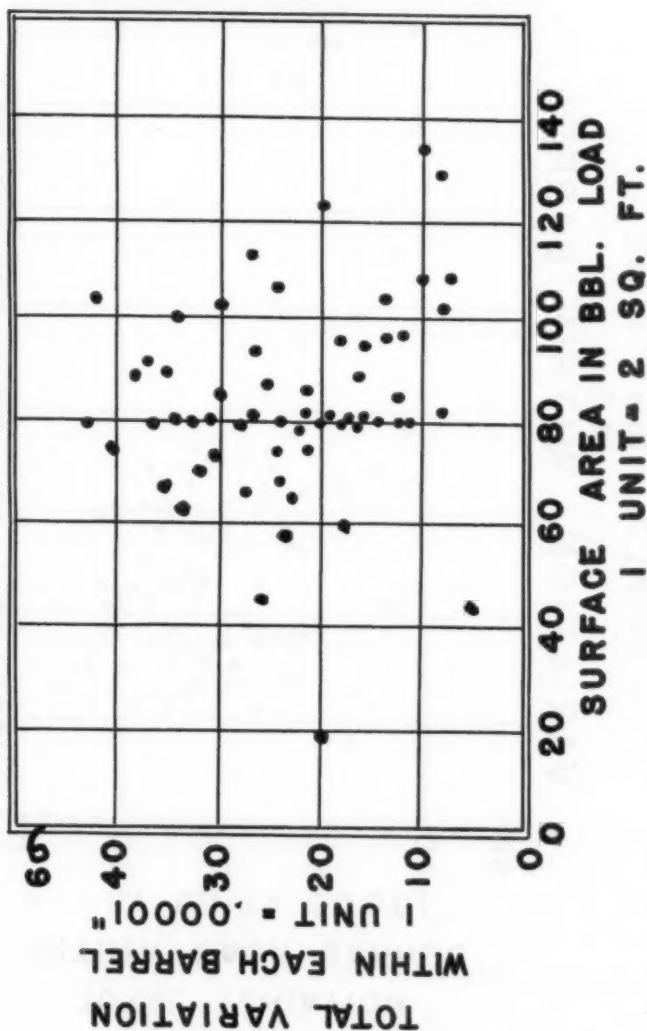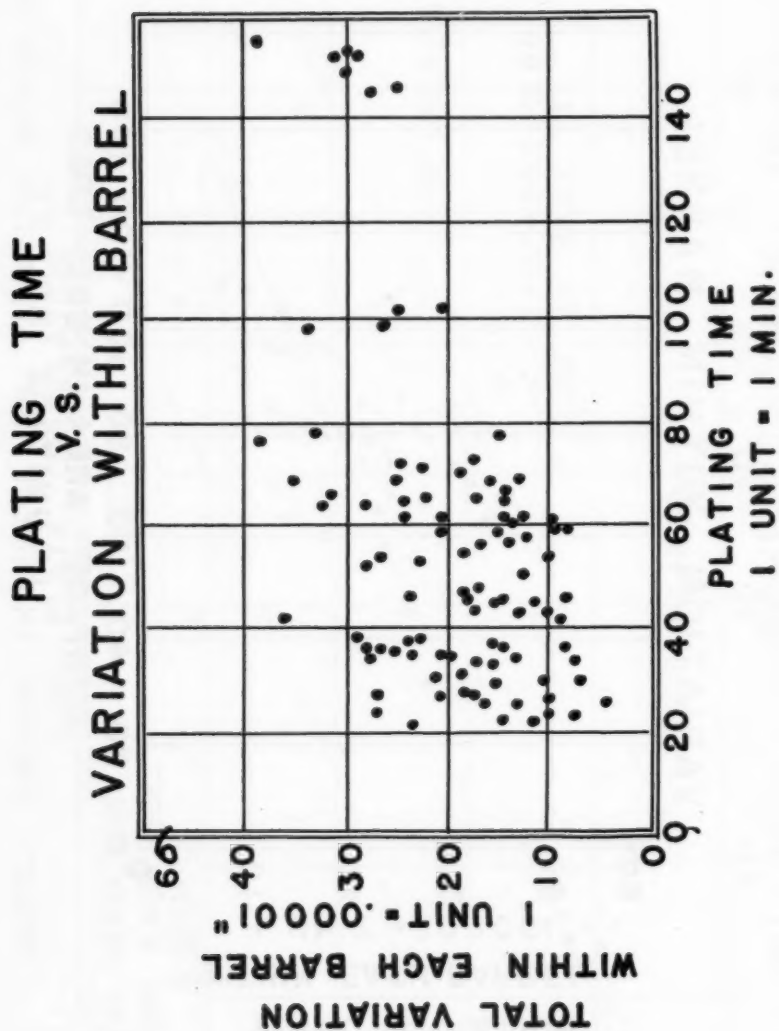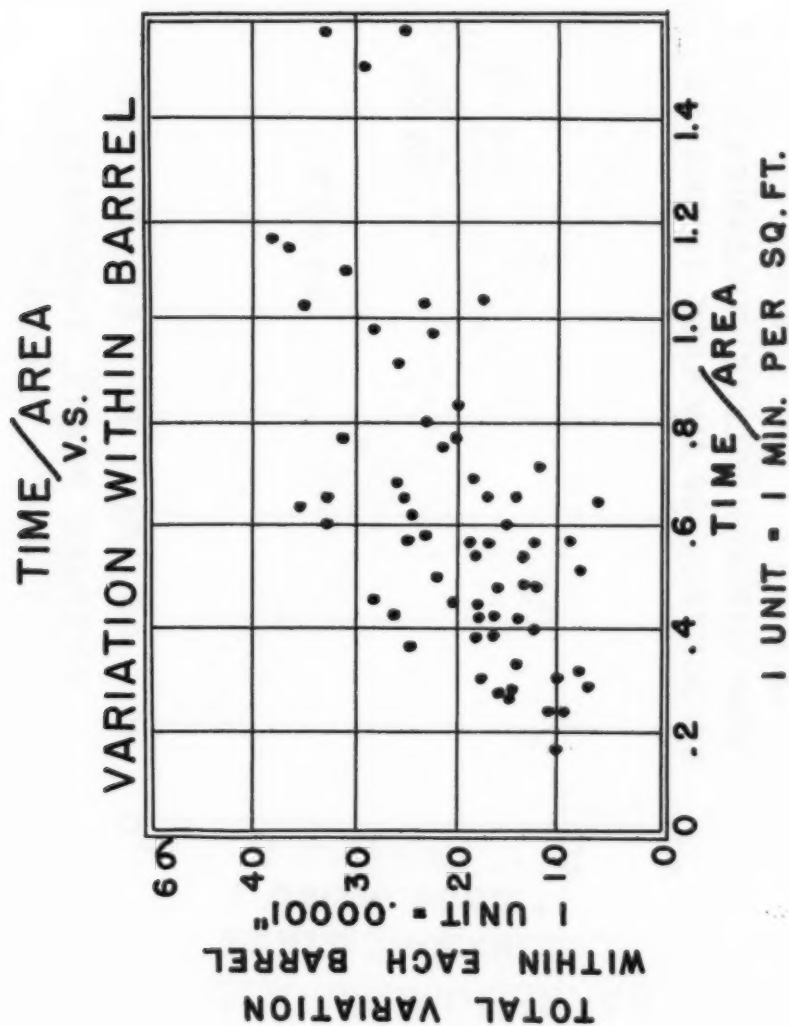CHART 4

345

QUALITY ASSURANCES
FOR MINIMUM PLATE THICKNESS
BY CONTROL OF AVERAGE THICKNESS

CHART 5

346
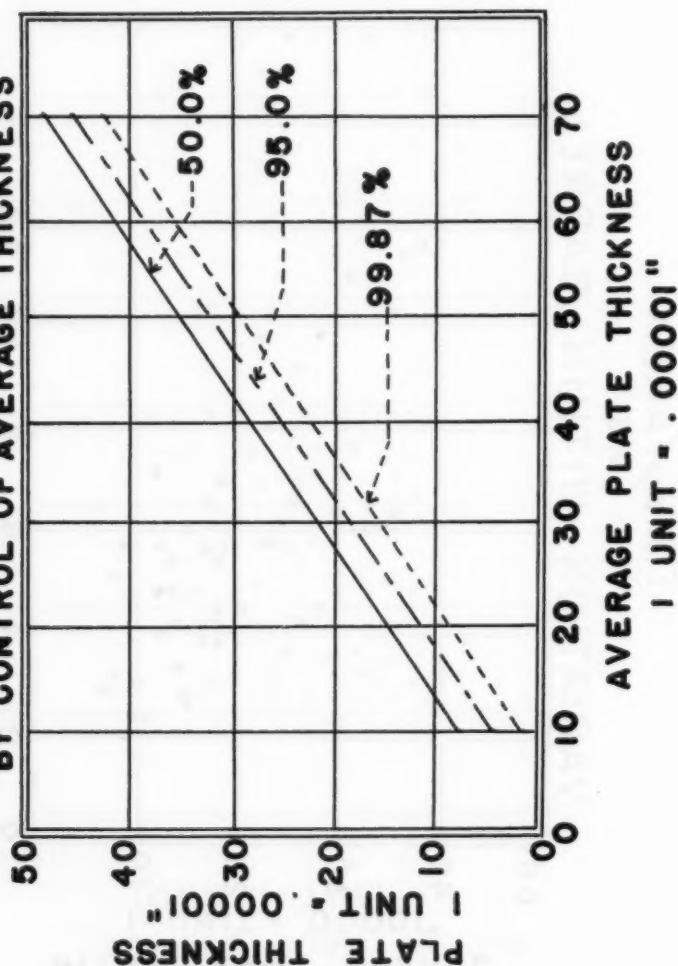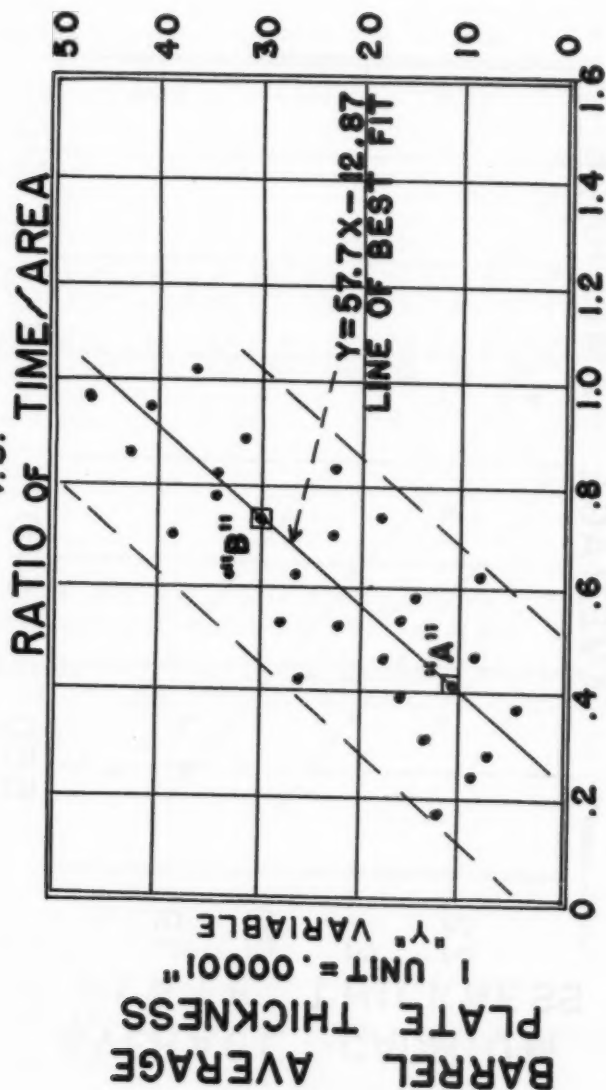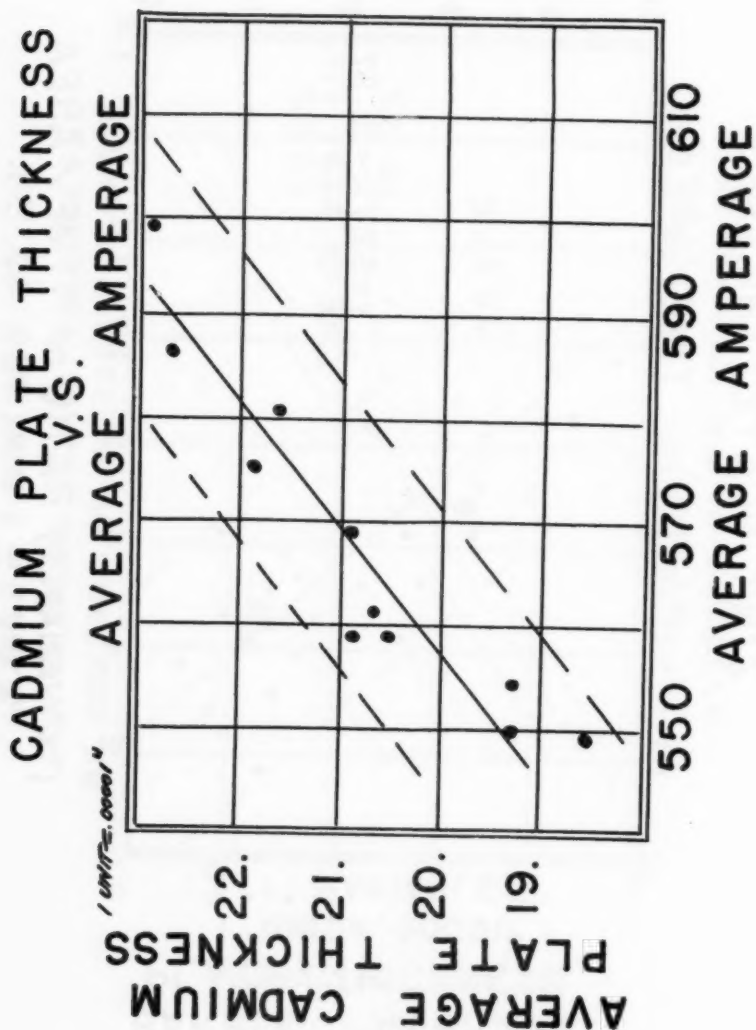
ZINC PLATE THICKNES
v.s.
RATIO OF TIME/AREA

$Y = 57.7X - 12.87$
LINE OF BEST FIT

("X" VARIABLE) RATIO OF TIME/AREA
I UNIT = I MIN. PER SQ. FT.

"Y" VARIABLE
I UNIT = .00001"
PLATE THICKNESS
BARREL AVERAGE

CHART 6

347

CHART 7

CADMIUM SOLUTION QUALITY CONTROL DATA SHEET	SPEC. & CHAR. 2.00 - 2.50 %	MACHINE "MACHINE" PART TANK

Sample #	1	2	3	4	5	6	7	8	9	10	11	12
10	FEB		MAR			APRIL		MAY		JUNE		
9												
8										"POSTED"		
7										Q.C.		
6										CHARTS		
5												
4										↓		
3												
2	2.34	2.51	2.73	3.14	3.38	3.34	3.95	2.84	2.55	2.63	2.55	2.26
1	2.67	2.17	2.06	2.79	3.16	3.30	3.22	2.90	2.75	2.63	2.60	2.40
Total												
Average	2.505	2.34	2.395	3.465	3.27	3.32	3.585	2.87	2.65	2.63	2.575	2.33
Range	.33	.34	.67	.65	.22	.04	.73	.06	.20	.00	.05	.14

Sample #	13	14	15	16	17	18	19	20	21	22	23	24
10	JULY			AUG.			SEPT			OCT		
2	2.38	2.19	2.16	2.17	2.21	2.20	2.25	2.13	2.17	2.16	2.17	2.16
1	2.41	2.21	2.25	2.15	2.14	2.20	2.23	2.21	2.15	2.19	2.20	2.16
Total												
Average	2.395	2.200	2.205	2.160	2.175	2.200	2.240	2.17	2.160	2.175	2.185	2.160
Range	.03	.02	.09	.02	.07	.00	.02	.08	.02	.03	.03	.00

Sample #	25	26	27	28	29	30	31	32	33	34	35	36
10	NOV.		DEC.									
2	2.21	2.27	2.15	2.19	2.16							
1	2.13	2.25	2.17	2.16	2.20							
Total												
Average	2.170	2.240	2.160	2.195	2.180							
Range	.08	.02	.02	.03	.04							

n	A_2	D_4	$3/d_2$	G_2	ΣX	$\bar{X}$	ΣR	$\bar{R}$	$A_2\bar{R}$	$D_4\bar{R}$	$3/d_2\bar{R}$	$G_2\bar{R}$
2	1.88	3.268	2.6596	.182	Control Limits based on Nos. 14-29							
3	1.023	2.574	1.772	.351	34.955	2.185	.57	.0256	.067	.116	.095	
4	.729	2.282	1.457	.401							$3\sigma_{\bar{x}}$	
5	.577	2.114	1.290	.423								
6	.483	2.004	1.184	.434								
7	.419	1.924	1.109	.441								
8	.373	1.864	1.054	.444								

CHART 8

349

QUALITY CONTROL CHART

"Quality Must Be Built Into Our Products."

Part Name	CADMIUM SOLUTION		Part No.		Mach. No.
Sample Code	ONE ANALYSIS WEEKLY (nz z)		Oper. No.		Dept. No.
Dim. & Tol.	MIN. 2.00 — MAX. 2.500		Char.		
Gage	PRIT EZ-h Two Consecutive WKS.		Mach. Name MACHINE " TANK		
Date	FEB. ————————————— SEPT. —— DEC.				
Time	WKly.				

1 Unit = **OUNCES PER GAL.** AVERAGES

$\bar{X} = 2.69$

$UCL_{\bar{x}} = 2.252$

$S_U \rightarrow$

$S_L \rightarrow$

$LCL_{\bar{x}} = 2.118$

POSTED Q.C. CHARTS

1 Unit = **OUNCES PER GAL.** RANGES

$UCL_R = .116$

GAGE — VISUAL — SCALE

Comments   **CHART 9**

FORM 112—MANAGEMENT CONTROLS, DES PLAINES, ILLINOIS

COPYRIGHT 1948

350

QUALITY STANDARDS IN SHOE MANUFACTURE

John T. Heald, Vice President and Factory Manager

When this country was first settled, there were amongst those earliest colonists shoemakers whose task it was to provide footwear for their communities. These men were craftsmen of the first order, and they accomplished their skills with little more than a few hand tools.

Today, more than 300 years later, we find the descendant of the original hand shoemaker still depending on his hands, eyes, and good judgment when he is called upon to make fine shoes for men and women. Whereas one might suppose that a considerable portion of the machinery used in the manufacture of shoes would be automatic, we find this not to be the case. Because of the great variations in the basic material used in shoes, namely leather, and the additional wide variations in the styling and construction of footwear, the shoe industry finds itself a far cry from automatic process machinery. The individual operator determines through his skill, in equal measure with the machine, the quality of the product. Thus, with so many present, it is of the utmost importance that some sort of definable fence be built around these variables to prevent them from running wild on occasion.

This becomes even more significant in the segment of the industry in which we, The Stetson Shoe Company, find ourselves competing - namely, high grade mens' shoes, retailing for about $25.00 per pair and up. It may be confidently and categorically stated concerning this area of the industry, be it mens' or womens' shoes, that the consumer expects a great deal of the product he purchases, as well he may. Were his standards of acceptance to be carefully analysed and segregated they would separate into the following:

- durability; wearing qualities
- comfort; fitting qualities
- style; eye appeal

To engender uniform and maximum customer acceptance and satisfaction has been the prime objective of our Quality Control Program since its inception three years ago. Again recognizing the tremendous variables present, which so often have been in the past responsible for variations in customer acceptance, we felt that the object of our efforts must be the attainment of Uniformity and Consistency of product.

It might aid to shed further light on this problem to define the components of our product quality:

(Chart A)

In order to assure maintaining the optimum level of customer acceptance in terms of these components it has been necessary to establish the following steps in analysing the product, a newly styled line being developed every six months and involving some one hundred new models or styles:

(Chart B)

351

In addition thereto constant vigilance is maintained over merchandise returned by customers for refund or repair. By this means another check is made of established standards designed to assure customer acceptance and satisfaction in the areas of material durability, foot comfort, and appealing design.

It would be fair to state, in passing, that the average person expects far more in terms of performance from his footwear than he does of any other item of apparel. Thus, the most exhaustive tests must be applied to the product to assure customer satisfaction; it is not at all unusual for customers to return shoes for dissatisfaction with one or more of the above quality components after months of obviously punishing wear.

Therefore, in an industry where it has been customary in the past for manufacturing personnel to carry quality standards about in their heads, permitting unhappy variations in the product to often occur between the time of product creation and its actual delivery to the customer, it became necessary for us to define these standards in order that everyone in the organization might understand and be in complete agreement upon them.

This has resulted in the development of a program which establishes standards for the following areas in the manufacture of the product:

- purchased materials quality
- process control standards
- finished shoe quality

Standards relating to purchased materials quality might be segregated into two groups:

1. Leather quality requirements, which pertain to material used in the upper portion of the shoe primarily, and which accounts for better than 50% of the material cost of the shoe, and

2. quality requirements for other materials, which include heels, welt (a basic ingredient of the goodyear welt construction, common to most mens' shoes) and other materials such as nails, threads, etc.

As concerns the latter category, tolerances as to texture and measurements of incoming materials have been established, suppliers have been advised of such, and samplings are made of each incoming lot. This has served to reduce considerably the reworking and repair of shoes in process which previously resulted from sub-standard materials.

The development and application of standards for upper leather materials, the major expense item of the shoe, will require further study before optimum controls may be considered to be in effect. This is due to the extraordinary variations in the animal skin itself, as well as the variations which occur between the several tanners in the grading of these skins. It can be reported that based on the fact that grading of the leather determines its price, a careful study of the relative cutting economy of the various grades of each tanner has indicated clearly that considerable savings can be effected, on a predetermined plan of purchase, by purchasing specific grades of leather for specific patterns or styles.

Continued refinement of this study of quality standards in grades of leather will result in very sizeable economies. Results in this area have been unsatisfactory in the past because of the lack of a carefully devised control plan.

Process control standards have been established with regard to critical operations, variations in which have in the past resulted in a finished product lacking in the desired uniformity, to the extent that a certain portion of these would be considered unacceptable and therefore rejected. Such in-process inspections or samplings are made daily, and unacceptable deviations reported to departmental foremen when detected, permitting corrective action to be taken on the spot, so to speak. In addition, weekly reports indicating quality performance in these critical areas are made to foremen and the plant manager each week, so that adverse trends in performance may be determined and the necessary action initiated. As such, the in-process checks provide an invaluable additional tool to the foremen and plant management in maintaining uniformity of the product.

Perhaps one of the most important areas over which standards have been established is that of Finished Shoe Quality. Herein a careful endeavor has been to define standards in terms of what the customer considers desirable and undesirable in the product, avoiding assiduously the pitfall of viewing the finished shoe quality from the viewpoint of the manufacturer rather than that of the consumer. Experience has indicated that an unintentional gulf can sometimes exist between these viewpoints. Thus defects are segregated as to severity, and are noted as either "major" or "minor". This differentiation has happily resulted in an appreciation that what may appear to the manufacturer as "major", may be construed by the customer to be "minor", and vice versa. The results of this audit, as is true of other samplings, are reported to manufacturing management, and in the process of taking necessary corrective action the customer viewpoint is properly balanced with and emphasized in relation to the manufacturers viewpoint.

It should be mentioned at this point that the establishment of these processes and finished product standards has served effectively to permit optimum control over the most elusive element with which apparel and footwear manufacturers in general must constantly deal - that of "style" or the artistic quality of the product. Sales and Production management work hard and long towards the objective of developing a product which has "eye appeal", the particular quality which the manufacturing department has the most difficulty in constantly and uniformly interpreting. In a so-called high grade and expensive product, the customer rightfully expects, (consciously or unconsciously) this quality feature to be pleasing to his taste.

Because of the unquestioned importance of finished shoe standards, only a slight deviation from the goal is permitted as acceptable, and results have indicated an ability to produce a finished product 98.6% free from major defects.

Since the desired results of any such Quality Control Program ultimately depend upon the understanding and response of all persons involved in the manufacturing process, management and labor alike, every attempt is made to keep such people constantly advised concerning the status of quality performance. Conferences are held periodically with manufac-

turning supervision, at which time performance is reviewed and plans are made to undertake effective corrective action. Quality Control best serves its purpose when the Foreman regards it as an effective tool and help in enabling him to best fulfill his quality responsibility.

In addition, a constant attempt is made to remind production employees of their responsibility in this area and the beneficial effect which proper observance of this responsibility has upon their job security. This phase of the Quality Control Department's activities, which can be designated as a quality-mindedness program, is effected through Quality posters, pay check messages, articles in a monthly newspaper, and employee plant inspection teams. The latter, which are comprised of selected employees from each department and their foreman, are briefed on the particular problems being encountered, and the significance thereof. They accompany the Quality Control Director on an inspection of the entire plant, and are asked to observe practices or conditions which may result in inferior quality performance. Over a period of time this serves to acquaint many employees with the relationship of their own tasks to those in other departments, and impresses them with the necessity of a high level of quality performance in the finished product.

In summary, the establishment of and control over standards in Shoe Manufacture has for this company greatly improved our ability to deliver a consistent and uniform product to the retail customer and hence to the consumer, an accomplishment which should help to inspire confidence on their part in the product and company.

It has also become evident that appreciable cost reduction may be realized through a Quality Control Program, despite the fact that the costs of maintaining quality have been historically low in relation to Sales Dollars. In addition to the savings resulting from reduction of in-process repairs and rejects, and customer returns, it is obvious that comparable economies may be enjoyed through purchase policies based on proven knowledge of the leather suppliers' grading methods.

In one of the country's oldest industries, steeped in tradition, quality standards have brought to this company an invaluable new vitality and confidence, which should continue to pay dividends in today's competitive market.

COMPONENTS OF PRODUCT QUALITY

Stetson Shoe Company

Basic Quality of Shoes	Quality of Design	Quality of Conformance
Pleasing Style	X	
Fine Materials	X	
Uniform Appearance		X
Comfortable Fit	X	X
Long-wearing Qualities	X	X

(Chart A)

STEPS FOR QUALITY ANALYSIS:

(Covering new products and selected old products)

1. Examination of materials for durability, work-
   ability, strength, color fastness, odor and
   uniformity.

2. Sewing tests on materials to determine optimum
   stitches per inch, together with evaluation of
   strength and appearance of sewed material.

3. Tests of increased severity on assembled shoes
   to determine points of weakest strength or
   potential damage.

4. Controlled wear testing of shoes with reports
   every 100 hours of use, on the basis of Form A.

(Chart B)

SYNTHETIC SAMPLING

A Way to Predict Circuit Reliability and to Automate Design

by

L. Hellerman and M. P. Racite

ABSTRACT

A new approach to the problem of designing reliability into an electronic circuit is the method of synthetic sampling. This method is superior to the commonly used "worst-case" design philosophy, which is characterized by the requirement that the circuit operate reliably when the components are at specified adverse limits (Reference 1). We shall point up this superiority by considering several shortcomings of the worst-case philosophy, as follows:

1. It makes no distinction between the probability of mean and extreme outputs.

2. If a circuit can fail in more than one way, no distinction is made between the probabilities of failure from the different causes.

3. It provides no measure of reliability.

Synthetic sampling does not have these weaknesses. It automatically weighs each output with its probability, and it gives the reliability of the circuit based on the actual component distributions. The method is illustrated in this report by application to a typical transistor switching circuit.

357

SYNTHETIC SAMPLING

A Way to Predict Circuit Reliability and to Automate Design

by L. Hellerman and M. P. Racite

INTRODUCTION

This paper presents an application of the method of synthetic sampling, or Monte Carlo, to the problems of reliable design and quality control of electronic circuits.

We believe a design procedure using synthetic sampling is far superior to the commonly used "worst-case" design philosophy, which is characterized by the requirement that the circuit operate reliably when the components are at specified adverse limits (Reference 1). We shall point up this superiority by considering several shortcomings of the worst-case philosophy, as follows:

1. It makes no distinction between the probability of mean and extreme outputs. In view of the nature of the output distribution, this is often unrealistic, unduly pessimistic, and costly.

2. If a circuit can fail in more than one way -- for example, from too high an output and too low an output -- no distinction is made between the probabilities of failure from the different causes. This unbalance in reliabilities, as an example will show, may actually make it desirable to design failures from one cause into a circuit purposely, in order to increase the overall reliability.

3. It provides no measure of reliability. The specification of limit component values for the worst case is somewhat arbitrary, and the designer never knows how much reliability he has bought by using one or another set of extreme component values.

Synthetic sampling does not have these weaknesses. It automatically weighs each output with its probability, and gives the circuit output distribution and reliability based on the component distributions used. An automatic design procedure using synthetic sampling on the IBM 704 will be described and applied to the design of a typical transistor switching circuit.

In the quality control of an electronic circuit a manufacturer may have the following problem. A certain lot of components to be used in the circuit is tested and found to be off specification. If these components are used anyway, will the effect on the quality of the finished circuits

be such that most will have to be discarded, or will the effect be marginal, so that most of the circuits will be satisfactory? Here again, as we shall see, the method of synthetic sampling is an excellent way to find the answer, for all we need to know is the circuit output distribution based on the off specification component distributions.

359

1. IMPROBABILITY OF EXTREME CASES

Design engineers have questioned the worst-case criterion on the ground that the occurrence of the worst case may be quite improbable, and the requirement that the circuit operate under these conditions may be too costly for the reliability that it buys. Statisticians may recognize the Central Limit Theorem in this phenomenon (Ref. 2). Our first object is to formulate a measure of it - some rule of thumb to back up the engineer's intuition.

Suppose we have a circuit whose performance parameter y is a function of several circuit parameters,

$$y = f(x_1, \ldots, x_n).$$

We assume the design engineer knows the distribution of each component. This information may be in terms of the distribution of the deviation of the component from its nominal value.

The deviation of the performance parameter is given approximately by the total differential formula,

$$dy = \left(\frac{\partial f}{\partial x_1}\right) dx_1 + \ldots + \left(\frac{\partial f}{\partial x_n}\right) dx_n. \tag{1}$$

Using the fact that the distribution p_i of

$$dy_i = \left(\frac{\partial f}{\partial x_i}\right) dx_i \tag{2}$$

is related to the distribution f_i of dx_i by

$$p_i(dy_i) = \frac{1}{\dfrac{\partial f}{\partial x_i}} f_i \left(\frac{dy_i}{\dfrac{\partial f}{\partial x_i}}\right) \tag{3}$$

the problem of estimating the likelihood of some deviation becomes the problem of estimating the sum of several random variables,

$$dy = dy_1 + \ldots dy_n. \tag{4}$$

This can be done by means of the composition of distributions formula of statistics:

$$p(z) = \int_{-\infty}^{+\infty} f(x) \, g(z - x) \, dx \qquad (5)$$

where $z = x + y$, f is the distribution of x, and g is the distribution of y.
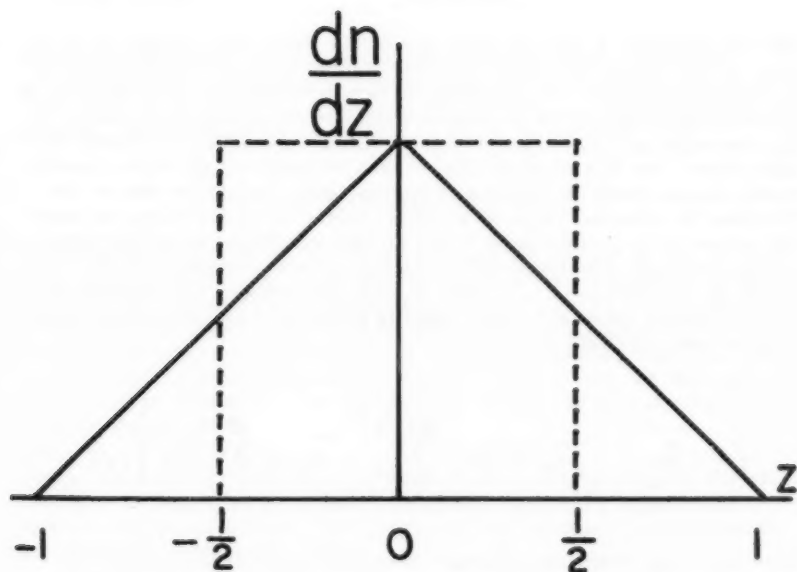

FIGURE 1. Distribution of $z = x + y$.

Fig. 1 illustrates the distribution of the sum of two random variables, each uniformly distributed from $-1/2$ to $+1/2$, as found by the above formula. The result agrees with intuition: The probability of the sum being near -1 or +1 is less than the probability of the sum being near 0.

If we now let f be the distribution we have just found for the sum of two variables, and reapply this formula, we get the distribution of a sum of three variables, and so on to n variables (see Appendix). The result of all this is the following:

> If the deviation of an output can be considered as a sum of n uniformly distributed and approximately equal random variables, then the probability of a deviation falling within an extreme nth of the total variation is roughly $1/n!$.

It is clear that this probability decreases as the number of components increases.

2. UNBALANCED RELIABILITIES

We now consider a second blind spot of the worst-case design philosophy: If the circuit has more than one output, there is no indication of a possible unbalance of the reliabilities with respect to each output. A circuit may be designed to operate when any particular combination of its elements are at the limits of their initial tolerances or end-of-life tolerances, but this does not insure that its mean output will be maximally distant from all regions of malfunction. Related to this is the fact that the nominal value of an output (obtained by inserting the nominal values of the components in the circuit equations) is not the same as the mean value of the output, intuition notwithstanding. On the contrary, if $y = f(x_1, \ldots, x_n)$ and x_i are independent, the relation between nominal value and mean value is given by a Taylor series whose first terms are (Ref. 4).

$$\bar{y} = f(\bar{x}_1, \ldots, \bar{x}_n) + \frac{1}{2} \sum_{k, l = 1}^{n} \sigma_l \, \sigma_k \left. \frac{\partial^2 f}{\partial x_l \partial x_k} \right|_{x = \bar{x}} + \ldots \quad (6)$$

Note that if the function is linear, mean and nominal values are the same.

The discrepancy between nominal and mean values is not merely of theoretical interest, confined to unusual circuits rarely encountered in practice. A close examination of the simple series circuit in Fig. 2 illustrates this difference.

Let $E = 1$ volt and $R = 1$ ohm, and suppose that R is uniformly distributed from $1/2$ to $3/2$. In this case the nominal value of I is 1, but the mean value is $\ln 3$. The probability density function is $1/I^2$, as shown in Fig. 3. (See Appendix II)

362
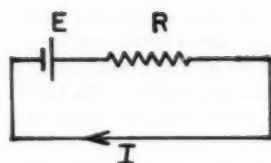
FIGURE 2. Simple
series circuit.
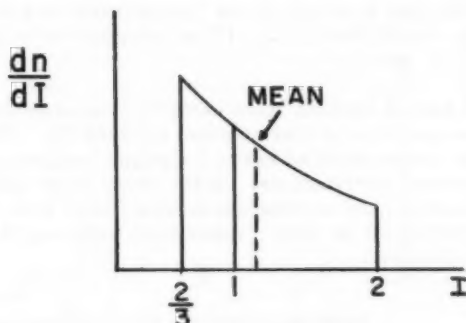
FIGURE 3. Unbalance of
reliabilities.

Thus the probability of a failure because of too small a current is
greater than the probability of failure from excess current. If a de-
sign engineer knew of this unbalance, he could compensate for it by
choosing R less than 1.

3. MEASURE OF RELIABILITY

A third difficulty with the worst-case design philosophy is that it pro-
vides no clear measure of reliability. An engineer using this method
will have the satisfaction of knowing he has considered reliability, but
he will have no measure of the effectiveness of this consideration.
This is really the key question of the entire reliability problem. To
answer it we need the output distribution. This in turn not only gives
a measure of reliability, but also tells us how the output is distributed
between its extremes.

There are several methods in the literature for estimating the distri-
bution of the output of a circuit. A rather popular one equates the
mean to the nominal value, and estimates the standard deviation of the
performance parameter y from

$$\sigma^2(y) = \sum_{i=1}^{n} \left[\frac{\partial f}{\partial x_i} \sigma(x_i) \right]^2 \tag{7}$$

Then, assuming a normal output, its probability density is given by

$$p(y) = \frac{1}{\sqrt{2\pi} \, \sigma} \exp - \frac{(y - \bar{y})^2}{2 \, \sigma^2} \tag{8}$$

363

This has been called the "propagation of error technique, " because of
the application of Eq. (7) to computation of precision measures. (Ref.
5, 6, 14.)

A second method, also using the linearity assumption, is based on the
composition of distributions formula (5). This has the advantage over
the propagation of error technique because we are not restricted to
normal distributions. In the study of the extremes of the distribution
function, the normal curve would have been quite unsuitable. Its tails,
trailing off to $\pm\infty$, would have submerged the very thing we wished
to observe.

THE METHOD OF SYNTHETIC SAMPLING

The method that we have found best for ease of application, adaptabili-
ty to available component data, and accuracy is Monte Carlo (Ref. 7).
In general, the essence of Monte Carlo methods is the conversion of a
problem to some equivalent statistical one and then solving this experi-
mentally, by observing a large number of cases. A classical example
is integration (Ref. 8). Our problem is a natural one for Monte Carlo,
for it is essentially statistical to begin with, and needs no conversion.
For this reason it has been called synthetic sampling.

Suppose we wish to obtain the output distribution of

$$y = f(x_1, \dots , x_n)$$

where the x_i are distributed in a known way from a minimum of x_{ia} to
a maximum of x_{iz}. We assume that the cumulative distribution func-
tion c_i of the i th component is such that, corresponding to any ran-
dom number between 0 and 1, there is a unique value of the parameter.

The flow chart for the synthetic sampling method is then shown in Fig.
4. The y_j's are a set of test points for the distribution. When the test-
ing is over, column 0 will contain all the cases less than y_0, column 1
will contain the number of cases between y_0 and y_1, and so on. Note
that the y_i need not be equally spaced, but can be chosen to highlight
any particular region of the distribution.

You may ask, what faith can we have in answers obtained in this way-
by chance, so to speak? How do we know that we can trust the random
number generator to be unbiased? The answer is that there are statis-
tical tests that can be and have been applied to these questions and that
tell us with what confidence we may believe the results. Studies of
this sort should be part of a proper application of the synthetic sampl-
ing method.

However, if possible, it is always comforting to test the method on a
problem for which we know the answer. For example, using the com-
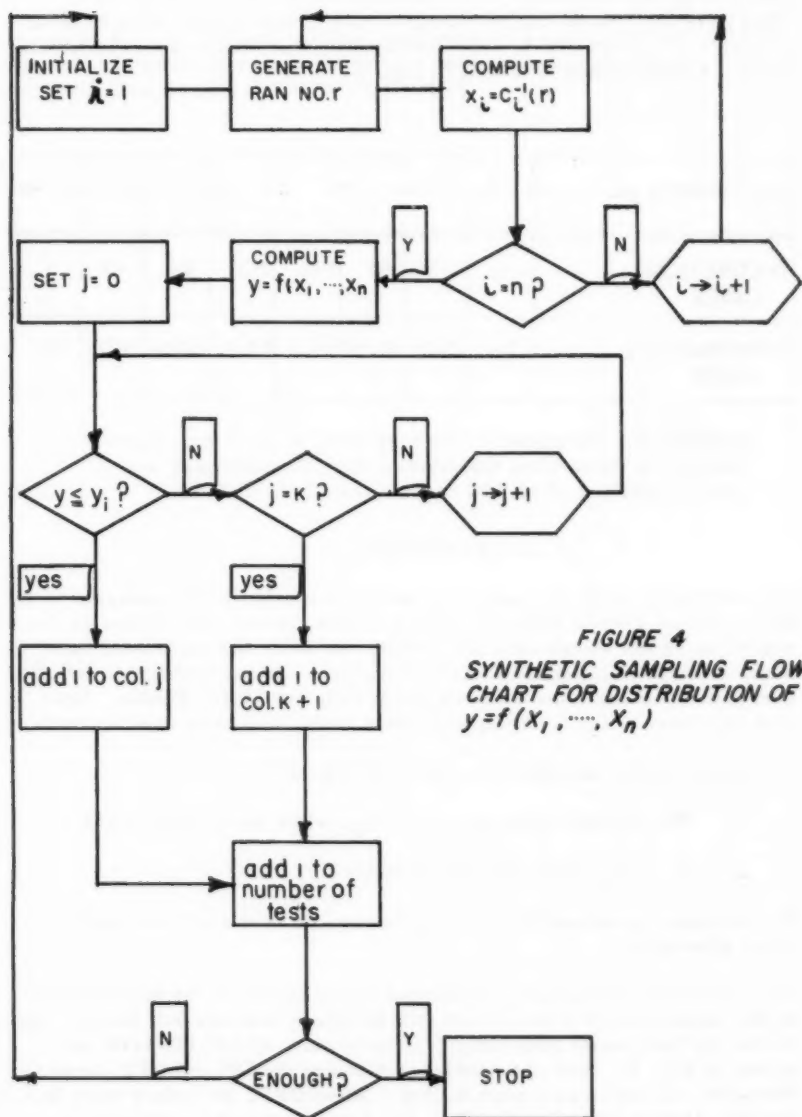
FIGURE 4
SYNTHETIC SAMPLING FLOW
CHART FOR DISTRIBUTION OF
$y = f(X_1, \ldots, X_n)$

position of distributions formula, we have computed the distribution of the sum of four variables, each uniformly distributed from 0 to 1. This gave the theoretical probabilities for z falling into each interval of length 1/2 from 0 to 4. The distribution was then found experimentally. A comparison is shown in Fig. 5.

TEST POINTS y_i	0	.5	1.0	1.5	2.0	2.5	3.0	3.5	4.0
THEORETICAL CASES	1	15	61	115	115	61	15	1	
EXPERIMENTAL CASES	.96	15.73	60.84	114.9	113.71	60.81	16.03	1.02	

FIGURE 5. Distribution of $z = x_1 + x_2 + x_3 + x_4$. Comparison of theoretical distribution for 384 cases with synthetic sampling of 38,400 cases normalized to 384.

EXAMPLE

The method of synthetic sampling has been applied to the analysis of the switching circuit shown in Fig. 6. This circuit was originally designed using the worst-case philosophy, in which the transistor parameters were at initial purchase tolerances. In our testing we let all parameter distributions c_i be uniform with end-of-life limits. Note that any other available component data could have been used as well.

Two performance parameters were considered:

1.   The voltage difference $V_e - V_1$, when the circuit is ON.

2. $V_e - V_b$, when the circuit is OFF.

For reliable operation these had to be greater than zero with some noise allowance.

Since we were particularly interested in the nature of the distribution in the neighborhood of zero - for this is where failures will occur - we chose our test points accordingly. The results of 100,000 tests are shown in Fig. 7. Note the unbalance between the ON and OFF cases. When ON, no case came near failing. When OFF, 693 cases were below the allowed noise level. This was a result which could not have been foreseen by simply considering worst cases.

In an effort to equalize the ON and OFF cases, the resistance R_3 was

366

decreased from its original design value of 2. 7k.   As R_3 decreased, the
ON case became slightly less reliable while the OFF case reliability
improved.   The optimum value appears to be between 2. 3k and 2. 4k, as
shown in Fig. 8.   With this value, and with the parameter distributions
we have assumed, the probability of failure from parameter drift is 1 in
10, 000.   That is, the reliability is 0. 9999.
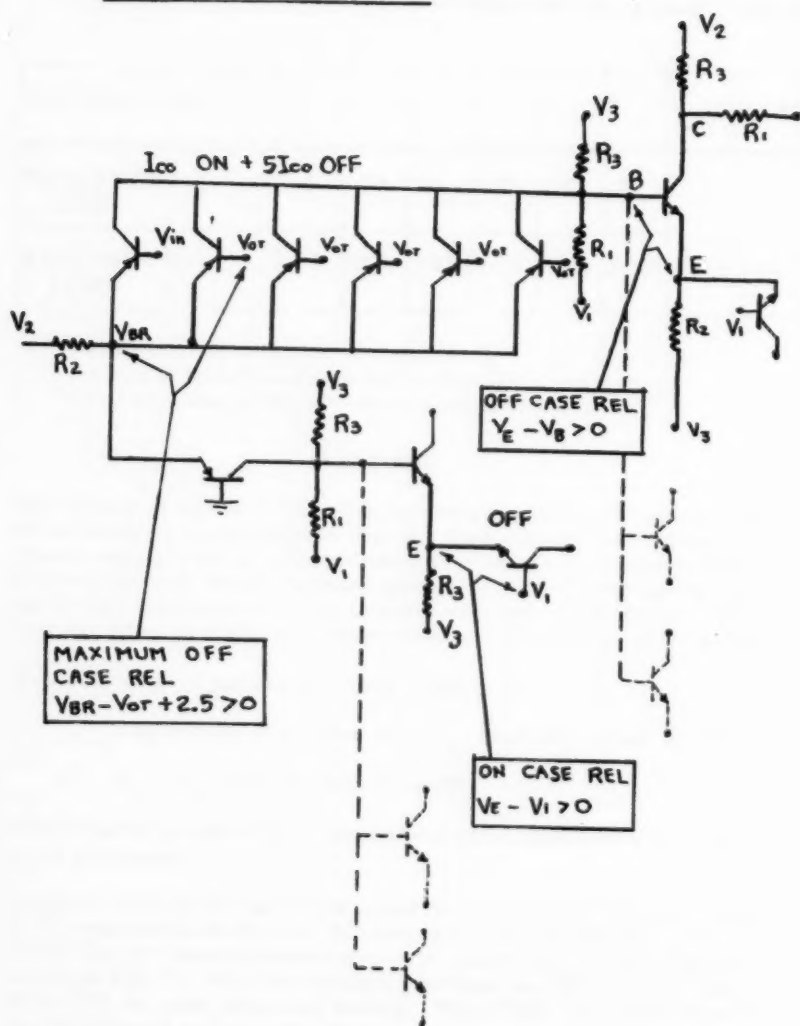
TRANSISTOR SWITCHING CIRCUIT

FIGURE 6

In the last section we were given a circuit design and component distributions and found the output distributions and reliability. By trying various values of R_3 we altered the output distributions and improved the reliability. We will now outline an automatic design procedure using the method of synthetic sampling.

Our design problem is the selection of the values for R_1, R_2, and R_3 in the switching circuit shown in Figure 6. Besides the two performance parameters considered previously for the ON and OFF reliabilities, we wish to consider the possibility of exceeding a maximum OFF bias. Thus we will have three performance parameters: ON CASE, OFF CASE, and MAXIMUM OFF CASE. Corresponding to each case there will be an output distribution which may be found by synthetic sampling following the flow chart shown in Figure 4.

Without loss of generality we may assume failures correspond to negative values of the outputs. (This can always be arranged by adding a suitable constant to the output, or by changing its sign.) Figure 9 then shows the three distributions for some starting resistance values. Our object is to find the resistance values that minimize the number of failures. Clearly, in order to do this we have to shift the worst case distribution to the right.

Our present program has a built-in intuition for selecting the parameter change to cause this distribution shift. For example, we know that the OFF case output is most effectively shifted right by decreasing R_3. Hence, if our program recognizes the OFF case as the worst case, it will decrease R_3 a small amount. This should improve the result and we may now test the distributions obtained with the new parameter values. Perhaps a different case will now be found worst. The program selects the parameter change to improve this, and so on, until the desired reliability is attained.

Various refinements suggest themselves. It is unnecessary to use Monte Carlo to test every change of the resistance values. Instead we may use the shift of the nominal value as a measure of the shift of an entire distribution. This saves time, since it replaces computation of the large number of samples necessary to find a distribution by computation of the single nominal sample. However, since nominal value shift is only a rough indication of distribution shift, we must "touch base" every so often with Monte Carlo tests. These serve as the basis for further changes. A flow chart for this procedure is shown in Figure 10.

Our program, shown in the flow chart has a further refinement. The more reliable the circuit becomes, the more current it will require from prior circuitry. Therefore, we want our design to minimize failures only within the capability of the circuit. This may entail shifting

some distributions left and others right in order that their left endpoints line up at some desired position. This position may be such that a predetermined number of failures will occur at end-of-life of the circuit.

Another refinement is an automatic means of selection of the parameter change for improving a particular case. The total differential formula

$$dz = \frac{\partial z}{\partial R_1} \ dR_1 \ + \ \frac{\partial z}{\partial R_2} \ dR_2 \ + \ \frac{\partial z}{\partial R_3} \ dR_3$$

may serve to make this selection. If we let dR_i be a fixed percentage of R_i the relative magnitudes of $(\partial z / \partial R_i) \ dR_i$ $(i = 1, 2, 3)$ give the effectiveness of R_i in altering z. The sign of the partial derivative tells whether an increase or decrease of R_i is needed to increase z.

IN CONCLUSION

To summarize, the worst-case design philosophy says: "Let the circuit operate when the circuit parameters have their extreme adverse values." In view of the shortcomings of this approach already discussed, we propose that it be replaced with the design philosophy which says, "Let the circuit operate with probability R when the parameters have their probable values." Synthetic sampling seems to be an excellent way to implement this philosophy.

370

FIGURE 7

371
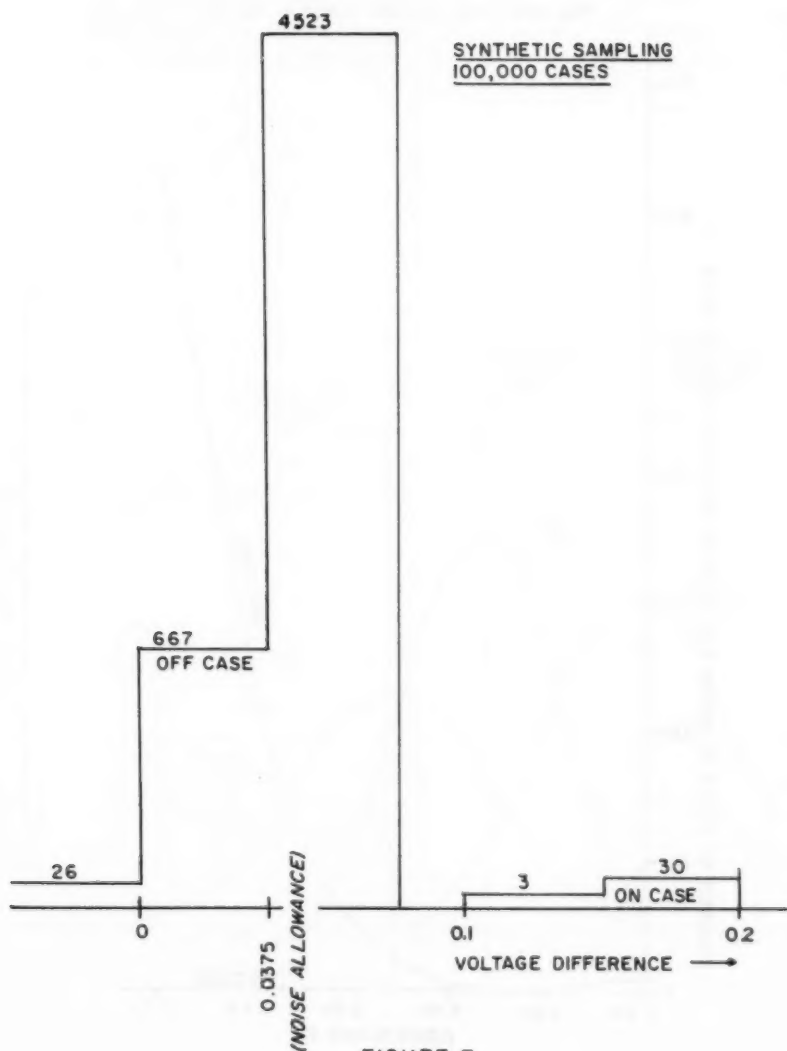
FIGURE 8

FAILURES IN 10,000 CASES vs. R₃

372
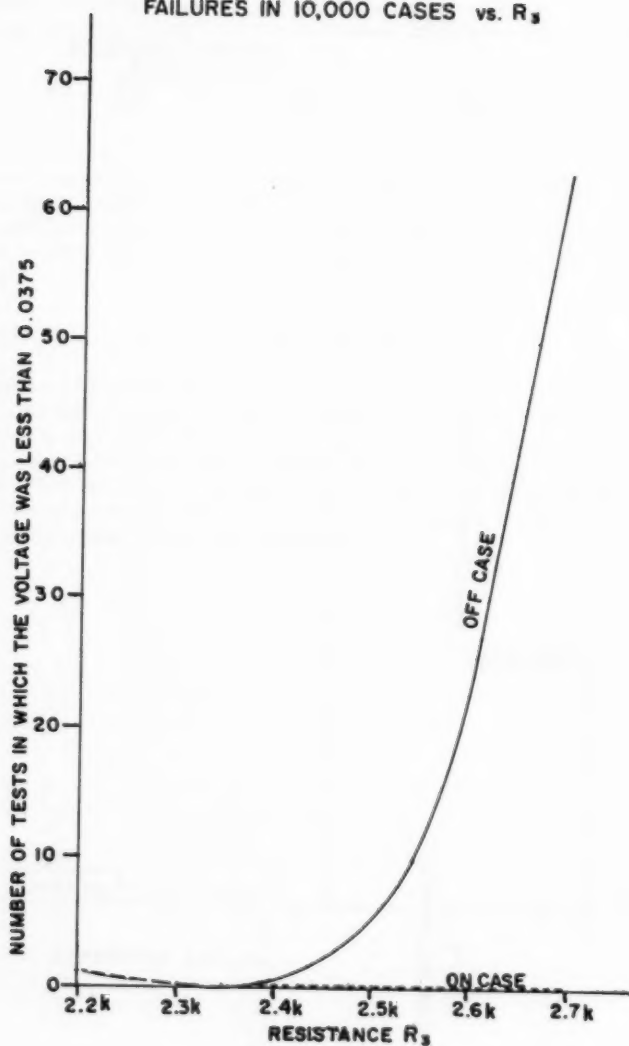
Monte Carlo output distributions of
ON, OFF, + BREAKDOWN cases of
switching circuit

Fig 9

Figure 10

374
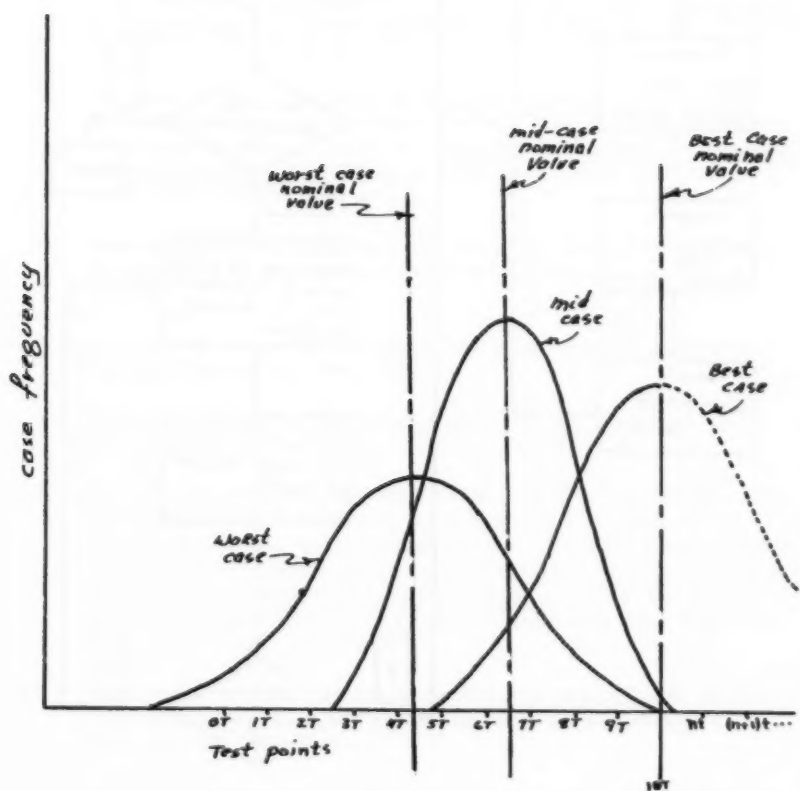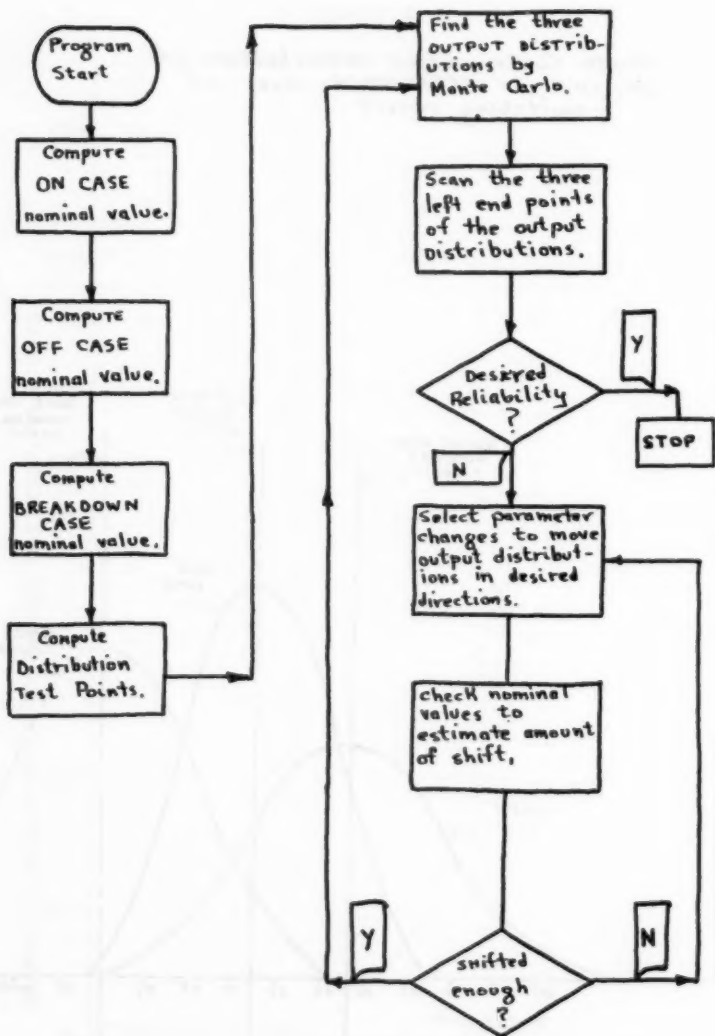
REFERENCES

1. Taylor, N. H., Designing for Reliability. Technical Report No. 102 (Unclassified), Massachusetts Institute of Technology, Lincoln Laboratory, December 9, 1955.

2. Cramer, H., The Elements of Probability Theory. John Wiley and Sons, Inc., New York, 1955.

3. Cramer, H., Random Variables and Probability Distributions. Cambridge University Press, 1937. Theorem 13, p. 36.

4. Meltzer, S. A., "Designing for Reliability." IRE Transactions of Professional Group on Reliability and Quality Control, September 1956, p. 36, formula (17).

5. Bartlett, S. B., The Method of Least Squares. Technology Branch of Harvard Cooperative Society, 1933. The formula for $\sigma(y)$ also appears in Hald, A., Statistical Theory With Engineering Applications, John Wiley & Sons, Inc., 1952, p. 118, and many other books on statistics.

6. Benner et al., "Designing Reliability into Electronic Circuits." Proceedings of the National Electronics Conference, Vol. X, 1954.

7. Brewer, E. E., "Mathematical Tools for Reliability Design." RETMA Symposium, May 1956.

8. Householder, A. S., Principles of Numerical Analysis. McGraw-Hill Book Co., Inc., 1953, Chapter 8.

9. The random number generator we have used is the 704-SHARE routine CL RAN 1 written by R. Johnson of the Lockheed Aircraft Corporation, California Division. The theory and analysis of this type of random number generator and other types is given in the Symposium of Monte Carlo Methods, held at the University of Florida, March 16 & 17, 1954. Published by John Wiley and Sons, Inc., 1956, New York.

10. The article by Butler in the Symposium of Monte Carlo Methods cited above discusses the generation of arbitrary distributions from uniform distributions.

11. Berz, F., "On the Cumulative Effect of Chance Deviations," Journal of the Royal Statistical Society, Series B, Vol. 16, 1954, pp. 269-284.

12. Parker, J. B., "The Accumulation of Chance Effects and the Gaussian Frequency Distribution." <u>Phil. Mag.</u>, 38, pp. 681 - 682, 1947.

13. Silberstein, L., "The Accumulation of Chance Effects and the Gaussian Frequency Distribution." <u>Phil. Mag.</u>, 35, pp. 395 - 404., 1944.

14. Schwartz, L. S., "Statistical Methods in the Design and Development of Electronic Systems." Proc. Inst. of Radio Engrs., N. Y. 36, pp. 664-670, May 1948.

APPENDIX I

Let us denote the sum of n statistically independent variables (each uniformly distributed from $-1/2$ to $+1/2$) by x_n, and the distribution of x_n by f_n. The composition of distributions formula, Eq. (5), then gives

$$f_2 (x_2) = \int_{-\infty}^{\infty} f_1 (x) \, g \, (x_2 - x) \, dx \qquad (9)$$

where $f_1 = g$, and $x_1 = x$. With the distributions we have chosen, the independent variables x and $x_2 - x$ are less than $1/2$ in absolute value. Thus

$$f_2 (x_2) = \int_{\max \left\{ x_2 - \frac{1}{2}, \frac{1}{2} \right\}}^{\min \left\{ x_2 + \frac{1}{2}, \frac{1}{2} \right\}} dx$$

and considerations of cases gives

$$f_2 (x_2) = \begin{cases} x_2 + 1 & -1 \leq x_2 \leq 0 \\ -x_2 + 1 & 0 \leq x_2 \leq +1 \end{cases}$$

We are interested in the nature of f_n in the neighborhood of its negative extremity, $-n/2$. Clearly, from Eq. (9),

$$f_2 (x_2) = \int_{-\frac{1}{2}}^{x_2 + \frac{1}{2}} dx \qquad -1 \leq x_2 \leq 0$$

377

Suppose

$$f_n - 1(x_{n-1}) = \int_{\frac{2-n}{2}}^{x_{n-1}+\frac{1}{2}} \int_{\frac{3-n}{2}}^{x_{n-2}+\frac{1}{2}} \cdots \int_{-1}^{x_3+\frac{1}{2}} \int_{-\frac{1}{2}}^{x_2+\frac{1}{2}} dx\, dx_2 \ldots dx_{n-1}$$

for $\quad \dfrac{1-n}{2} \leqq x_{n-1} \leqq \dfrac{3-n}{2}$

Then, by Eq. (5).

$$f_n(x_n) = \int_{-\infty}^{\infty} f_{n-1}(x)\, g(x_n - x)\, dx$$

where f_{n-1} and g are zero outside the ranges

$$x_n - \frac{1}{2} \leqq x \leqq x_n + \frac{1}{2}$$

and

$$\frac{1-n}{2} \leqq x \leqq \frac{3-n}{2},$$

Hence

$$f_n(x_n) = \int_{\max\left\{\frac{1-n}{2},\ x_n - \frac{1}{2}\right\}}^{\min\left\{x_n + \frac{1}{2},\ \frac{3-n}{2}\right\}} f_{n-1}(x)\, dx$$

For $\quad -\dfrac{n}{2} \leqq x_n \leqq -\dfrac{n}{2} + 1$, this gives

378

$$f_n(x_n) = \int_{\frac{1-n}{2}}^{x_n + \frac{1}{2}} \int_{\frac{2-n}{2}}^{x_{n-1} + \frac{1}{2}} \cdots \int_{-1}^{x_3 + \frac{1}{2}} \int_{-\frac{1}{2}}^{x_2 + \frac{1}{2}} dx \, dx_2 \ldots dx_n$$

for $\quad -\dfrac{n}{2} \leqq x_n \leqq \dfrac{2-n}{2}$

Differentiation of this gives

$$f_n'(x_n) = f_{n-1}\left(x_{n-1} + \frac{1}{2}\right)$$

and for the (n-2) th derivative we have

$$f_n^{(n-2)}(x) = f_{n-1}^{(n-3)}\left(x + \frac{1}{2}\right) = \ldots = f_3^{(1)}\left(x + \frac{n-3}{2}\right) = f_2\left(x + \frac{n-2}{2}\right)$$

Since $f_2(-1) = 0$, it follows that $f_n^{(n-2)}\left(-\dfrac{n}{2}\right) = 0$, similarly for the lower order derivatives at $-\dfrac{n}{2}$. Also since

$$f_n^{(n-1)}(x) = f_2'\left(x + \frac{n-1}{2}\right) = 1 \qquad \text{at} \ \ x = -\frac{n}{2}$$

we may infer that the order of contact of $f_n(x)$ at the point $-\dfrac{n}{2}$, with the x-axis, is n-2. Furthermore, since f_n is a polynomial of degree n-1, the nth order and higher derivatives are all zero. All this says

$$f_n(x) = \frac{\left(x + \frac{n}{2}\right)^{n-1}}{(n-1)!} \qquad \text{for} \ \ -\frac{n}{2} \leqq x \leqq \frac{2-n}{2},$$

379

and the probability of falling in this region is

$$\frac{1}{(n-1)!} \int_{-\frac{n}{2}}^{\frac{2-n}{2}} \left(x + \frac{n}{2}\right)^{n-1} dx = \frac{1}{n!} \, ,$$

which is the result we wished to prove.

Alternative methods of arriving at the same result are given by Silberstein (Ref. 13) and Parker (Ref. 12). The article by Berz (Ref. 11) is also of interest in this connection.

APPENDIX II

Let the cumulative distribution of R be P(R). Since R is assumed uniformly distributed from 1/2 to 3/2 we have

$$P(R) = \int_{\frac{1}{2}}^{R} 1 \ dr = R - \frac{1}{2}, \qquad \frac{1}{2} \leqq R \leqq \frac{3}{2}$$

The probability that the current is less than or equal to I is the same as the probability that R is greater than $\frac{1}{I}$, because of the inverse relation between these variables. Hence, the cumulative distribution of I is

$$P(I) = P \left\{ R > \frac{1}{I} \right\} = 1 - P(R) = \frac{3}{2} - R = \frac{3}{2} - \frac{1}{I}$$

The probability density of I is

$$\frac{dP}{dI} = \frac{1}{I^2}$$

The mean of I is

$$\overline{I} = \int_{\frac{2}{3}}^{2} I \ \frac{1}{I^2} \ dI = \ln 3$$

382

CONSUMER RESEARCH AND QUALITY CONTROL

Robert Weller
Alfred Politz Research, Inc.

CONSUMER RESEARCH OBJECTIVES

The objective of consumer research should be to enable the marketer to build his sales economically at the rate at which he buys consumer research. To accomplish this goal, consumer research must have certain characteristics.

1. It must cover all aspects of marketing -- product, distribution, merchandising, advertising, pricing, etc. Advertising research cannot be a separate entity and separate from consumer product research, and still be productive. All phases of marketing are too interrelated.

2. It must be creative -- not just informative and entertaining. Productive research today even gives direction to management regarding the kinds of products to produce in advance of production. Conducting this pre-product research does not mean discovering what consumers want, but rather what consumers will accept. In analyzing the field of marketing, it doesn't take much research to find out that the obligation of the marketer is not to give the consumer merely what the latter wants. The consumer wants everything at the lowest possible price, and, at the moment, the consumer has no imagination, not even for his own benefit. The consumer depends upon the creativeness, inventiveness, and pioneering of industry, and the marketer has to come in and make the consumer want what has been invented and what the marketer desires he should want. That's what is called selling, and that's the nutshell of all marketing operations.

Suppose we had conducted typical "find-out-what-he-wants" surveys in 1800 and asked consumers to tell us what to do to improve their products, conveniences, etc. If such opinions had been gathered and followed, after these meetings all of us would ride back to our various offices in the finest horse-buggys you ever saw, reading, under the most efficient kerosene lamps, newspapers that contain news five days old.

This does not mean that we do not study consumer's wants. We study them to help us learn how to create new wants and desires.

3. It must be predictive -- To be creative, consumer research must be able to predict. Prediction involves risks, including the risk that the possible inadequacy of the research will eventually be detected. Some think this risk can be avoided if research just confines itself strictly to descriptive research. This concept, however, is both selfish and illusionary; it does not change the principal issue. It only suggests a division of labor in which the professional researcher is careful not to participate in drawing productive conclusions from the statistical data he collected. The conclusions, and therefore, the predictions, are then left for somebody else to make -- the sales executive, the manufacturer, the company research director, or management in general. The point still remains that consumer research is not useful unless it leads to predictions.

Guiding management means providing a basis for making decisions. A decision entails the expectation of a specific result; therefore a decision entails a prediction. Worthwhile research arrives at the causal conclusion that action A will lead to result B. A decision can then be made to take action A, with the knowledge that result B will occur. Here is where valid research can increase tremendously management's knowledge level and lower management's risk. Of course, whether action A ought to be taken or whether result B is worth the extra effort is not the problem of the researcher, but of management. This is so because management not only must consider the consumer angle, but also company policy and philosophy. Often these are of a higher order than simply the consumer aspect.

4. It must use experimental design -- To be predictive, consumer research must be conceived and executed as an experiment. The survey is but our laboratory in which we put to test our ideas of our products, our advertising and our merchandising. It's where we test reactions -- reactions of consumers in an effort to learn more about their attitudes, their motives and their behavior. Being conceived of as an experiment, consumer research then must start with an hypothesis, it must use or create a mechanism for testing the hypothesis and then must draw a conclusion about the acceptance or rejection of the hypothesis. The research task then is an exercise in experimental design, and the design of experiments requires great creativity and imagination. The consumer researcher cannot hope to have statistics or psychology or some other discipline do this creative work for him. Engineers make use of various sciences, but it is the existence of specific design problems which makes engineering a science of it's own. Consumer research, too, and predictive research in particular, has it's own unique design problems which should make it a science of it's own.

THE PRODUCT OF CONSUMER RESEARCH

Having these characteristics, what then is the "product" consumer research offers to management? The product is basically "information." Information in the form of recommendations of suggested management action, and in the form of conclusions based upon the data collected. The data collected are, in a sense, created by the researcher. They are created because they are reactions to stimuli we put to the consumer. The researcher creates the questionnaire. But what validity can we attach to these data? In short, how do we control their quality?

THE FACTORY THAT PRODUCES CONSUMER RESEARCH DATA

The factory that produces the data for consumer research is, of course, the survey itself. There are many "processes" and "machines" that help produce a survey.

First, is the design of the sample and the actual carrying out of this design in the field. The utilization of probability sampling has given consumer research the foundation and framework within which it can operate close to the level of the pure sciences. Without such a framework, validity could not be insured, and reliability statements could never be made. Alfred Politz introduced probability sampling to the commercial field and has been a major influence for it's growth among consumer research organizations.

384

A second part in the survey production is the questionnaire itself, which is perhaps the most important piece of machinery in our factory. Third is the field -- interviewing. Fourth is editing the questionnaires and coding the questionnaires for translation to machine tabulation; fifth the mechanical translation of the information to punch cards or tapes and sixth the machine tabulating. Seventh is the computions and eighth is the analysis which includes recommendations for management action. Each of these parts can be adapted to quality control techniques.

The relative importance of each of these parts and how they are balanced in reference to their cost and their contribution to the overall quality of a survey, is discussed in the Appendix of a study we conducted for Life Magazine, called "A Study of Four Media." It is contained in the chapter called "Balanced Accuracy."

Today I shall limit my remarks to just one of the eight steps above, "Field Interviewing." In this discussion, we shall assume that we have a properly executed probability sample and that our questionnaire is a properly functioning piece of machinery. The questionnaire is put into this working order by adequate pre-testing before the survey is launched.

The importance of field work has been discussed and examined many times in literature. There have been several investigations that have actually quantified the contribution that interviewer variation makes to the total mean square error in a survey. One was a paper given by Steve Stock, called "The Analysis of Interviewer Variance In Block Samples." This study was conducted in a single city, with blocks used as primary sampling units. Of the total variance, the major portion was contributed by interviewer variation, and equalled approximately 58%.

ADMINISTRATIVE CONSIDERATIONS

The most important contribution to the assurance that quality work will come from the interviewing staff is to have top-notch personnel. This, of course, is a very obvious statement, but it nevertheless needs to be said over and over again. The development and build-up of a field interviewing staff does not mean to accumulate a list of names. It does mean to establish a close relationship with interviewers, treating them as employees of the firm and gaining their loyalty. This everyone takes for granted, but it is not at all easy to develop such a relationship. It can be developed only if the research company can provide interviewers with a sufficient volume of work. This does not mean current volume as much as it means a promise of work in the future. Work guaranty and security are important to interviewers; more important than high-paid spotty work. In our company, in order to further solidify this relationship, we have put a good proportion of our interviewers on a minimum wage guarrantee basis, regardless of whether they work or not. In return they agree to work exclusively for us.

To build a field organization in this manner, it is immediately obvious that the total number of interviewers a company can have is directly related to the volume of business it has. Our company averages about 6,000 interviews a week, or one-third of a million a year.

Even with this tremendous volume of interviews, our regular staff consists of only 600 interviewers. We do have a few hundred more whom we can pretty well depend upon, which we turn to during peak weeks when we're at the 3 sigma level instead of the average.

So long as the total number of interviewers is directly related to the dollar volume of a company, so must be the total number of primary sampling units a company uses. Six years ago we were using 55 primary sampling units; three years ago 103; today we are moving up to 152. This expansion in primary units corresponds to our dollar volume expansion. Had we, five or six years ago, tried to use 152 primary sampling units, we would have been using names and not quality interviewers in these areas. In a sense, this is a case of the "rich getting richer." But it still is axiomatic. Only large organizations can use a large number of primary sampling units with a large interviewing staff if high-quality field work is expected.

INTERVIEWER TRAINING AND SUPERVISORS

A common fallacy about consumer surveys is that interviewers have to be personally trained for each survey. With a solid field organization, such a procedure is highly inefficient and wasteful. Interviewers are no different from other field personnel, such as salesmen or missionary men. Preliminary training is needed, but continual training is not. For very special and unusual surveys, training may be necessary. An example of this is our recently completed "Life's Study of Consumer Expenditures." This was such a gigantic study, plus the fact that new field and questioning techniques were used, that training was a must. This kind of training, however, is the exception and not the rule. Moreover, with a loyal field staff that has been with an organization for some time, interviewers will soon accumulate more experience than we in the home office, and consequently, will know more about interviewing than we do. This training fallacy grew because research organizations had names of interviewers, not a quality field staff. The remedy was, therefore, put forth that interviewers must be personally trained for each survey -- a needless and wasteful effort.

While interviewers need not be repeatedly trained, occasional personal supervision is necessary. We call our supervisors, "Regional Instructors" because their principal functions concern recruitment of interviewers and trouble shooting. Their supervisory duties are limited. Each instructor is assigned one or more primary sampling units. Altogether, we have 69 of them. All are on a minimum wage guaranty; all have been seen by members of our Field Department, and almost all have been to our New York offices.

AREAS FOR MEASURING INTERVIEWER PERFORMANCE

In spite of having a top-notch field organization which we know from years of experience performs excellently, it is still necessary to check the quality of every survey produced being sure each meets performance standards.

In a typical consumer survey we must measure interviewer performance in the following areas:

1. Do they actually ask the questions? (The cheater problem.)
2. Do they omit questions?
3. How efficiently do they complete their assignments? (cost per interview)
4. How accurately do they make observations?
5. Do they misunderstand questions and thereby mix questions up?
6. Do they distort questions?

Depending upon the performance areas we're checking and the control program established, different units of measurement and sample sizes are used. A unit might be a single question, a group of questions, an entire questionnaire or the entire work of an interviewer.

PROCEDURES FOR MEASURING PERFORMANCE

The first performance area, "The Cheater Problem," can be handled in three ways:
1. Post card checks
2. Telephone checks
3. Personal re-interviews

It is rarely necessary to re-visit respondents in person in order to verify whether or not they have been interviewed. Post card or telephone checks are all that is needed. In making a post card check, in addition to determining whether or not an interviewer called, a few other questions selected at random from the questionnaire are asked again. This provides us with several points to check for consistency.

It has been our experience that 15-20 interviews per interviewer is sufficient to ascertain whether or not the interviewer has conducted his interviews. In selecting a sample, we select a few interviews at random from each of his assigned locations. On the average, a location consists of about five or six interviews. In a typical post card check, then, we will select three interviews from five to seven of the locations for each interviewer. We do not select more interviews per location because if an interviewer has cheated, it usually means he has not visited the location at all and a single interview would be sufficient. We do not select less than three per location, however, because we do not get 100% response to our cards, and we want to get at least one return per location. The response rate averages about 40% to 45% on post card checks. Actually, it is much more likely for a person who has not been interviewed to return the card than for a person who has been interviewed. We have set up the following acceptance and rejection procedure:

If any post card is returned because of "no such address" or "no such party at address," a 50% check on the remaining interviews of the interviewer is sent out. If we receive two such returns, a 100% check is made on his remaining interviews. If any card is received with the note that he was not interviewed, a 100% check is made on the remaining interviews. If there is an inconsistency concerning a particular question re-asked, a 50% check is made on the interviewer's remaining work. If two inconsistencies show up, a 100% post card check is made. From the second mailings, if any discrepancies at all are ascertained among the returns, the entire work is then held aside for a further check by a Regional Instructor. All invalid work is

discarded.

Telephone checks are used if a study is either a local one or one in which there is an extremely short field schedule, with interviewing having to be completed within one or two weeks. The procedures and criteria used are identical with the post card check.

In reality, complete cheating is a very rare occurrence. A more common problem is the partial or hurried interview, whereby the interviewer will either skip many questions, asking only the key ones, or so hurry the interview that poor quality results. This is more difficult to detect than cheating, and it is here where quality control techniques are most valuable.

The second performance area, omitting questions, is a simple one for establishing standards and for quickly ascertaining whether or not an interviewer is under control. Here the unit of measurement is an entire location. Out of the total work in a location, the percentage of questions omitted is determined. Our standard is 5% with appropriate limits depending upon the total number of questions. The first three locations completed by the interviewer are used. If an interviewer is found to be out of control by omitting too many questions, he is immediately contacted and told of his errors and his entire work on that survey is then checked for omissions. An interviewer is said to be under control on omissions if his first three locations are in control. His remaining work on that survey is then not further checked for omissions.

Cost per interview, performance area 3, is of great concern in any survey. We usually have past studies of sufficient similarity so that we can use standards already developed. In any single survey, however, we use two standards; one for metropolitan areas and one for non-metropolitan areas. While costs for reaching respondents is higher in the non-metropolitan areas, actual interviewing time is higher in metropolitan areas. The first four locations are selected per interviewer and the average cost per interview per location is computed.

Interviewers who fall outside of the limits are immediately examined. In this cost area, however, more often than not, those falling outside the limits can usually be explained by unusual circumstances. In probability sampling, anything is liable to happen simply because interviewers often find themselves in "unusual" places. They have to detour because a bridge has been washed away, or the mud may be so heavy as to cause them to walk long distances or hire a tow truck to pull them out of the mire. Unexpected interruptions during the interview, poor weather, etc., all can cause costs to skyrocket. For those interviewers for which no assignable cause can be discovered, two actions are taken. First, if his costs are between two and three sigmas, a letter is written to him, giving the average costs for work similar to his and suggesting how he might improve his performance, and the rest of his work for that survey is examined. If his costs are beyond the three sigma limit, he is asked for an explanation and a detailed check of his past performances is made, and his remaining work checked. If these checks, and/or his explanation are not satisfactory, the interviewer is dropped. In such cases, he is usually inefficient or is simply making excessive charges. If an interviewer's first four locations are under control, remaining checks are not made.

388

For these first three performance areas, standards from past surveys can usually be used. For the last three, concerned with accuracy of observations, understanding of questions and the distortion of questions, past surveys are of little help, with the exception of the case where we have repeat surveys. In all tailor-made or "first-shot" surveys, we need to develop standards from the current survey itself. This seriously limits the application of quality control techniques.

The most obvious way to establish a criterion for evaluating the quality on such custom surveys is to have more experienced personnel "re-do" the interview, or parts of the interview. It is usually done by a Regional Instructor and assumes, of course, that he gets the "correct" answer. This call-back procedure is very expensive. In addition to it's expense, the great disadvantage is that it is time-consuming. Unless there is ample time for field work, the call-back method cannot be used. Moreover, certain questions do not lend themselves to such treatment. You cannot ask them again and hope to get the same results. Knowledge questions, attitude and opinion questions, usually have to be avoided in a call-back procedure. Also, unstructured questions (or so-called "depth" questions) fall in this group, for with such questions considerable latitude is permitted the interviewer. To try to duplicate the stimuli is virtually impossible.

The main advantage with the recall method is it's simplicity, for acceptance and rejection criteria can be readily established. Sol Dutka, President of Audits & Surveys, has used this method with excellent results. He does not put all interviewers under such control in each survey, but rather rotates them, controlling only a portion on each study. This method is described in The American Marketing Proceedings of the Detroit Meetings, 1957.

If past standards and re-interviews are not feasible, then each interviewer's work must be compared to the work of all other interviewers. There are two major problems with this procedure. The first problem is that differences between interviewers may not be because of performance, but because of real differences between one interviewer's group of respondents compared to other interviewer's respondents. For example, if an interviewer is assigned only rural locations, or only high-income locations, etc., we can expect this response to be different from others.

The remedy would be to randomize locations among interviewers. This, however, leads to excessive travel costs. From a pure cost point of view, it is, of course, far more efficient to assign locations as close to an interviewer's home as possible. But, this interferes with out quality control. A compromise is usually effected. In the assignment of locations by the Field Department, both accessibility and variety of areas are balanced subjectively. In this way, interviewers are assigned a variety of areas without being too expensive to get to.

The second major problem facing this technique is the inadequate sample size for each interviewer. We are now not operating with an "agree or disagree" situation as we are with the repeat interview technique. We are now concerned with more or less random variation. Our standard is something like: "26% of the public gives as it's

389

reason for no longer using product Y, poor absorption." The total
work load an interviewer will have in a survey averages between 40-60
interviews. Even if we used his entire assignment, we would be using
very wide tolerances, too wide to be sensitive enough to be used for
improving quality.

    To overcome this difficulty, we are experimenting with what
we call, "Pattern Sheets." This is a principle which simply uses the
relationship between questions as a criterion, thereby increasing the
sensitivity and discriminating power of the test. These sheets are at
present a visual device, and we have not as yet quantified their
results. Several questions are placed across the top of the sheet and
several placed down the side. A mass of results from questionnaires
are plotted to establish the general relationships. The sheet will
now look much like a scatter diagram. Several copies of these sheets
will then be made up. On each copy, several interviewer's answers will
now be plotted, using different colors and symbols for each interviewer.
When an interviewer is out of control, a distinct pattern takes shape
that is readily visible by following the same colors and symbols. These
Pattern Sheets are very effective in detecting interviewers who do a
partial interview and then fill out the rest of the questionnaire at
home. This system works because interviewers, like everyone else, do
not act as the laws of chance, nor have they the marketing superiority
to predict what relationships will show up. Consequently, patterns
show up very quickly.

    The following chart is an illustration of a portion of a
Pattern Sheet. Down the left we have listed a question pertaining to
frequency of use of one of our client's products, brand Y. Across the
top we have two other questions; one pertaining to attitude toward
price of brand Y, the other, frequency of use of brand X. There is no
overall relationship between price of Y and attitude toward Y.
However, notice that Interviewer A shows up with a strong relationship
while Interviewer B was in control. The other part of the illustration
shows a good relationship between usage of brand Y and it's largest
competitor, X. Here the same interviewer (A) wrongly guessed that no
relationship would exist. Interviewer B was again in control.

    We believe these sheets show great promise as they are not
nearly as dependent on sample size nor on complete randomization of
location, for relationships like this are much less dependent upon
location characteristics than they are on behavioral characteristics,
which are much more randomly distributed.

PATTERN SHEET ILLUSTRATION

391

Edward W. Armstrong
Simonds Saw & Steel Company

The basic procedures of Quality Control can be used in an existing manufacturing operation to improve the level of quality, increase production and at the same time reduce the cost of manufacture of many products. This paper is presented to illustrate how a typical application is made and what economic rewards can sometimes be realized.

Let us compose a synthetic situation and say that we are in the business of producing irregular shaped cutters, that our cost is satisfactory, the quality of our outgoing product is good and our customers are happy. Our management, however, would like to see a lower ratio of indirect to direct labor. The conditions are as follows:

> Production   - 100,000 cutters per week
> Factory Cost   - Approximately $.36 each

Thus we are allowed:

> 100,000 x $.36 x 50 weeks = $1,800,000
> per year to purchase raw stock, direct
> labor and all the various burden items
> connected with producing our product.

Accordingly we decide to experiment with Quality Control, knowing that most of the characteristics of our irregular cutters are visual and that our job will not be easy. Our personnel are all responsible for quality as well as quantity. Our inspection consists of a large group of final inspectors and we do not have roving inspectors. However, we assign one man as a Q. C. inspector in our roughing area. The first operation on cutters is rough forming. Our inspector, working with a relatively small sample from each lot and reporting what he finds on a form as shown in Figure 1, tells us that each operator is producing some defective pieces and that a few of the men contribute much more than their fair share of rejects. We soon find that we need a sample board to guide our inspector in his decisions so, with the aid of our supervisors, one is developed.

Our inspector classes the type of defect which he observes as; Wrong location, A; Dull tool, B and Twist, C. We soon find that either A or B are operator controlled and we discuss them with supervisors and operators to encourage improvement in their work. Type C occurs only about 2% of the time in some lots but we can't seem to stop it. Finally we decide that it occurs mostly on one machine and is not limited to any particular size of cutter or operating shift. We ask "WHY?" One operator suggests that he knows his forming machine does not always stop in the same place and possibly that is the cause of our trouble. The master mechanic agrees that it is possible so the machine is repaired and C defect is reduced. Our supervision could not previously control the problem because it occurred so infrequently. Rather, they were accustomed to living with it.

			QUALITY CONTROL DATA SHEET — ROUGH FORM CUTTERS			

SIZE _____ DATE FORMED _____

LOT NO. _____ DATE INSPECTED _____

MACHINE NO. _____ INSPECTOR _____

OPERATOR NO. _____ LOT SIZE _____

NUMBER INSPECTED	NUMBER DEFECTS	FRACTION DEFECTIVE $\overline{p}$	WRONG LOCATION A	DULL TOOL B	TWIST C	NOTES

FIGURE 1

Working with the information taken from the Q.C. inspectors reports, we make fraction defective charts (p charts) of each operator for a few days, without showing them to the men. A sample chart of our fastest man, who has a history of producing at 150% of a normal rate, is shown in Figure 2. It is apparent from the first part of the chart that he also produces far too many defects and initially his $\overline{p}$ is 10.2%. Then we explain the charts individually to each man. We state that, if the machine does not run right, if the tools are not right, if the steel is not right, we will fix it. We tell them our belief that the work can be improved. We ask them to tell us if they can't do it right. We hang the charts on the machine.

Immediately we encounter from the operators a rash of reasons for defective work. We proceed to eliminate the reasons by repairing equipment, revising tooling, assisting operators to perfect their technique and doing whatever else is necessary. Within a few days, our efforts; combined with the psychological effect of the charts, has encouraged each man to produce better work. Our offending (best) operator has made a great change in the quality of his work. His $\overline{p}$ is now 1.2%. He has discovered that, because all equipment, tools and raw stock are controlled better, he can still produce at his usual very fast rate and not make many defects.

After another three weeks, our "best" operator is still puzzling to us. Many of the other men are doing better work than he does, yet we know him to be highly skilled. We discuss the reasons with him. Suddenly he says "When I form the last piece, it is sometimes too short." I can't locate it right. Yet, if I bury it in the good pieces, I get paid for it. I won't do that again." His $\overline{p}$ is .3% from then on.

394

PERCENT DEFECTIVE CHART
ROUGH FORM OPERATION
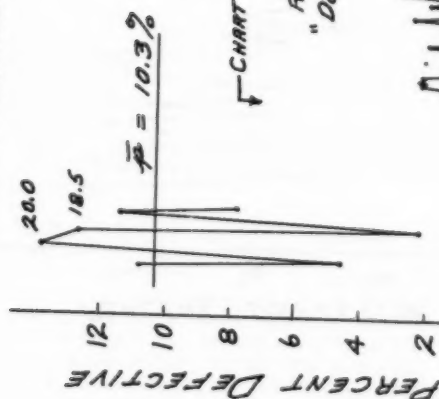DAILY INSPECTION — CUTTERS
OPERATOR 18-128
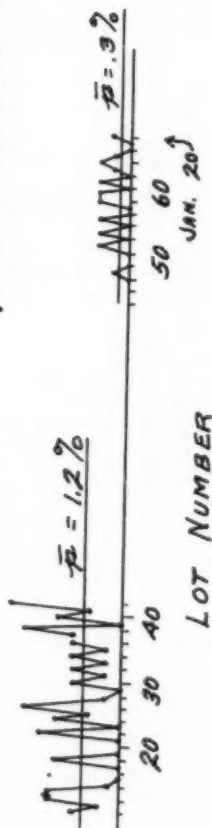
FIGURE 2

PERCENT DEFECTIVE CHART
ROUGH FORM CUTTERS
DAILY INSPECTION – ALL OPERATORS

INCLUDES 1020 C

PERCENT DEFECTIVE
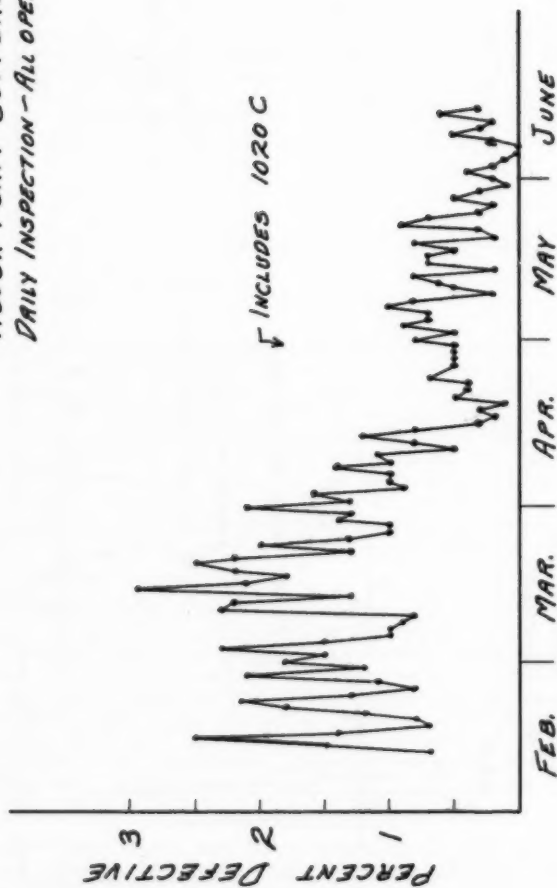
FIGURE 3

A "p" chart for the entire rough form department is made as shown in Figure 3. It indicates that, within a few months, we have saved at least 2% of the total production as good cutters rather than defectives. We previously found the defectives at final inspection. A rough calculation of the dollars saved in the forming area is as follows:

SAVING IN FORMING

100,000 Cutters per week x 2% x $.36 each = $720.00 per week

$720 x 50 weeks   = $36,000.00 per year

In the meantime we have had policy decisions to make. We have been asked, "Who screens these defective lots?" And we have answered, "You made them, you find them." We have created quality consciousness among all the people who make rough cutters.

All of the other operations in the manufacture of cutters naturally come under our consideration. We install scrap cans at each operation and ask everyone who finds defective pieces to put them in the can. Periodically we count the amount in the cans and determine where our rejects are occurring. We are like the F.B.I. after it's 10 most wanted men because we always have some operations that are much worse than others. We emphasize the fact that a continuous chain conveyer is continually causing many defects although it only comes to our attention when the conveyer jams and we must shut down for a while. We convince management that it is to their advantage to replace the conveyer. Soon a different type conveyer is installed and our periodic problem has disappeared.

The periodic count of rejects from each operation also emphasizes that Final Inspection can be a source of great information to us. We decide to use that source to our best advantage. We have several inspectors doing what is called 100% inspection. We provide them with a form so they can report how many cutters they look at each day and what they find wrong. Immediately, we see the need of training each one to work to a definite level of quality. A sample board, showing the various possible defects with Go and Not Go examples, is therefore developed. The record from the inspectors is accumulated for a few weeks and a fraction defective chart as shown in Figure 4 is made. We find that the level of quality of cutters incoming for inspection is 9.6% defective. A measure of outgoing quality is also made so that we will have a reference later when we need it.

A "Gold In The Mine" survey, made to justify our request for a full time Q. C. investigator, shows that 80% of the defects are due to one operation called "Final Sizing." It is obvious that we must work on this operation. We ask one inspector to report only the defects due to final sizing and to classify these according to the reason for their occurrence. We find that the reason most frequently assigned is "Under Adjustment." It is controlled by the operators and shows a tendency to leave extra stock for Re-Sizing if necessary. A fraction defective chart for Final Sizing is made as shown in Figure 5. We note that it closely follows Final Inspection, as it should.

397

PERCENT DEFECTIVE CHART
CUTTERS — ALL SIZES
FINAL INSPECTION

$\bar{p} = 9.6\%$

MERRY X-MAS
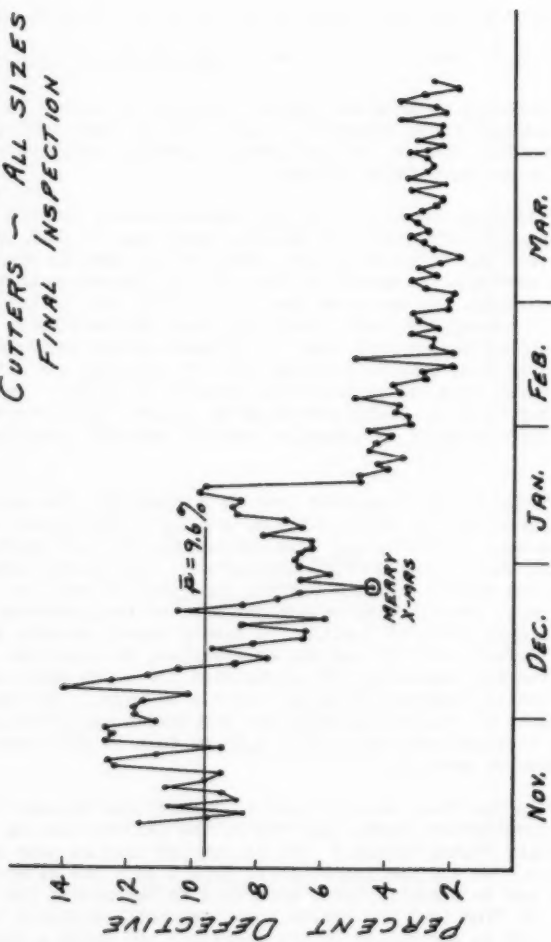
FIGURE 4

PERCENT DEFECTIVE CHART
CUTTERS — ALL SIZES
FINAL SIZING OPERATION
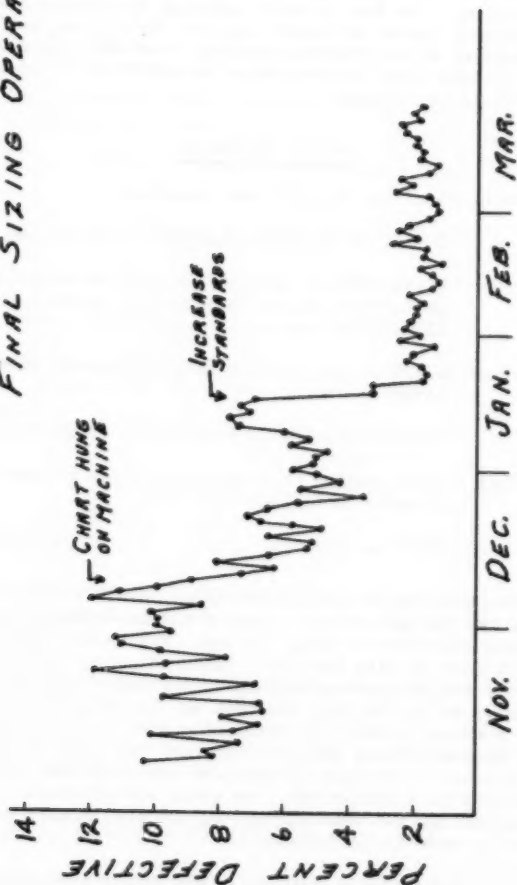
FIGURE 5

All of the defects due to Final Sizing are returned there for rework. A liberal time standard is allowed for the rework because it must be done in small lots. Half of the cutters are eventually salvaged and half scrapped. We prepare Fraction Defective charts of each operator on Final Sizing and hang them on the machines. We note some improvement but not enough. We conclude that our operators seem to like the rework. Experimentally, we determine that it is possible to operate a Final Sizing machine and produce nearly all good pieces if the incoming work is satisfactory. The operator must, however, keep the setup adjusted to the mean specification. We make a small increase in allowed time to cover the small necessary amount of rework and ask that it be done immediately rather than sending it to final inspection. Our fraction defective charts soon drop to a level which is considered satisfactory. An estimate of the saving made is as follows:

SAVING IN REWORK

About 8% of the Cutters were reworked.

Rework Direct Labor amounts to $78.00 per week

Re-inspection Labor (at a slow rate because individual gaging is required) amounts to $120.00 per week.

About 3-1/2% of the Cutters were ultimately scrapped after rework.

100,000 x 3-1/2% x $.36   = $1260.00 per week

Thus a total saving from the elimination of rework amounts to $1458.00 per week.

$1458 x 50 weeks   = $72,900 per year

Naturally, when the incoming work is not satisfactory, we must investigate and try to improve it. One of our investigations leads back to the previous operation of Rough Sizing. Try as we will, this roughing operation continues to give trouble. Since we cannot control it, we are led back to the preliminary operations to try and obtain more control there. Finally we decide that the best way to control Rough Sizing is to perform a previous operation differently and eliminate the Rough Size completely. When new tools are provided, we find that it is possible so the change is made. A battery of machines are discarded and their operators are assigned to a better job. We could not eliminate the operation before because we did not have control of the preliminary steps. An estimate of saving made is as follows:

SAVING BY ELIMINATING UNNECESSARY OPERATION

The Direct Labor required for Rough Sizing amounts to $670.00 per week.

$670 x 50 weeks   = $33,500 per year

400

As we continue to watch the level of quality incoming for Final Inspection, we observe that most of the lots are very clean. It is easy to prove that our outgoing quality level with a good plan will be better than it was when we checked after 100% inspection before Q.C. was installed. A sampling plan is started so that we can find these good lots quickly and not waste time on them. Inspection thus becomes a job of determining the level of quality. It can be done much faster. More than half of our inspectors are reassigned to production work. A saving as follows is made:

SAVING BY SAMPLING INSPECTION

The inspection labor (Indirect) saved amounts to $574.00 per week.

$574 x 50 weeks = $28,700 per year

SUMMARY OF SAVINGS

Saving in Forming Area.........$36,000 per year

Saving in Rework.............. $72,900   "

Saving by Elimination of
   Unnecessary Operation.......$33,500   "

Saving by Sampling Inspection..$28,700   "

Total Saving.....$171,000 Per Year

Conclusion:

We are allowed $1,800,000 to buy material, labor and burden. We have saved practically 10% of that amount.

In addition, our ratio of indirect to direct labor is satisfactory.

Our product is better. We make it that way. Our customers confirm it.

Our product is less expensive. We made a substantial saving in manufacturing cost.

Fortune Magazine called Quality Control "the sharpest management tool to come down the pike in 50 years." Sometimes it seems that way!

401

WHAT DOES THE AIR FORCE EXPECT OF CONTRACTORS

Colonel J. G. Schneider
Chief, Quality Control
Hq Air Materiel Command

Introduction

Gentlemen - It is a pleasure to discuss the subject, "What Does the Air Force Expect of Contractors," because it provides an opportunity for me to follow-up on two previous papers presented to the American Society for Quality Control.

In 1951 at Cleveland, I introduced to the Society members their first look at the Air Force's new quality assurance program explained by a new Specification MIL-Q-5923 (USAF). This document was our formal answer to the question, "What does the Air Force expect of contractors?" I might add that the reception afforded my discussion of this new concept, as those of you who were there might remember, was like most things new, not too enthusiastically received.

In Philadelphia in 1953, "The Philosophy Underlying the Specification MIL-Q-5923" was the subject of my talk. I might add that after two years trial, the acceptance was far more favorable than the original introduction, and now upon returning to Quality Control, after four years, the many complimentary things I hear about 5923 from contractors' representatives are most gratifying. In our procurement activities overseas we tell the contractor what we expect of him through the medium of 5923. It may interest you to know that within the few years since it was originated, this specification, or facsimilies thereof, have received wide acceptance and adoption by our military counterparts and industries in many friendly and allied nations throughout the world. It is presently written in the languages of most of the European NATO countries, as well as Great Britain, Canada and Japan, where it is furthering the use of quality control as a management tool.

Top Management and Quality

When speaking of "contractor" I am not only directing my remarks to the quality managers within industry but to top-level management as a whole. I believe we can accept as a basic principle that the quality of any product is largely dependent upon the policy of management, and the effort, interest and emphasis which all of management devotes to the quality and reliability of its products.

By "management" I mean those members of the organization who have responsibility and authority to plan, program, direct and control the industrial effort.

The existence of a quality control organization can be nothing more than a paperwork exercise and so much "window dressing" unless its efforts are firmly backed and supported by all other elements of the organization, and is not being subordinated to any interests which could negate quality.

Effective quality control within the industrial organization be-
comes a means to an end which can be summed up in terms of a quality
product, shipped on schedule, to customer satisfaction and, last but
not least, at a reasonable profit. If we accept these principles,
then we must look next to the quality control manager and his obliga-
tions which can be summed up in a few brief words, "Assure that the
quality and reliability are in accord with established standards in
the most economical way practicable with a minimum effect on the in-
dustrial process." This is an over-simplified statement of a job in-
volving a great multitude of diverse activities which must be planned,
coordinated and intermeshed with practically all functions and opera-
tions of the whole industrial enterprise.

Quality Control and Reliability

With the spectacular development in technology and the complexity
of present day weapons systems, and with particular reference to mis-
siles, which can be "performance tested" only on a sample basis, some
rather unique developments have taken place. The emphasis on relia-
bility in the midst of the complexities of unmanned missiles, and
space vehicles emphasizes the need for professional quality control
engineers, new, better, and less costly inspection techniques, and
statistical methods which far outstrip yesterday's activities. We
learned in attempting to define reliability[1] and then to measure it
that there is a need for closer relationship between all industrial
activities than ever before. Quality and reliability take on new and
added significance in the missile era.

Today, quality control and reliability planning must be coinci-
dent with the initial proposal stage and simultaneously follow thru
the development of models and test articles, the prototype and finally
the production article. Modern techniques must be incorporated into
the plan and properly integrated with the overall engineering, buying,
production planning, manufacturing and testing activities. All func-
tions and procedures for controlling quality and reliability must be
reduced to written instructions and provisions made for documentation
of the quality control effort. The system must be designed to prevent
occurrence or recurrence of defects and a means for positive timely
corrective action.

The Contractor's Scope of Responsibility

The contractor and his quality manager can no longer regard
"quality control" as an "in plant" operation. His scope of respon-
sibility and concern extends itself in one direction, into the sources
and suppliers of a great variety of critical components and parts, and
in the other direction, into the vast complex of the military esta-
blishment, the depots, the test centers, the operating commands, and
even to bases scattered world-wide.

To further compound the situation, there are many instances where
the design contractor will develop, manufacture, test, support, main-
tain, assist in the training of military personnel, and construct and
operate special facilities for a particular weapons system.

404

Within all these areas of activity, quality control plays a vital part as a key element of top management. In fact, it is management in the broadest and most exacting sense, involving a high level of professional, scientific, technical and managerial skills. Planning the quality control program is as important and precise as all other planning for a given project.

Of significant import is the need for data which gives a concise measure of quality control effectiveness and the quality of performance and reliability during production, test and after delivery to the customer. Quality Control should be so positioned within the organization to assure unbiased analysis of data concerning quality and reliability and to redirect action information to other responsible elements of the organization in a fast, positive and complete manner. To be really effective, there must exist a closely integrated relationship among and between all participants in the industrial program. There is an ever increasing need for verifiable evidence of positive control of quality of each item, component and subsystem as it is passed from vendor to subcontractor to prime contractor.

Vendor Rating System

One very effective technique in accomplishing this is the use of the vendor rating system. This involves the use of preaward surveys, as well as a post award running evaluation of vendor's quality effectiveness. It also includes the use of data feedback to the vendor and establishes a factual vendor quality rating which may be utilized for future business. I note that quite a number of contractors are utilizing such a system in some degree. The development of a well planned vendor rating system is not only an effective control device but permits more economic utilization of quality control manpower. This, of course, is in consonance with the Air Force Quality Control Specification 5923. This specification requires the contractor to have an effective and economic quality control system, and does include positive quality assurance of subcontract or vendor parts. The use of a vendor rating system, which requires that evidence of inspection, test and control of quality is furnished by the vendor or subcontractor with his product, can result in minimizing costly visits or assignment to the subcontractor's plant. I firmly believe that the subcontractor should provide proof of quality to the prime just as we look to the prime to show proof of quality to the Government, whenever he tenders materiel for acceptance.

Cost Consciousness

In the short period since returning from overseas, I have attended a number of meetings with various industry associations. In almost every instance, I have been impressed with the fact that cost of quality and cost of quality control are generally interwoven in most presentations. On 1 March, the Dayton Section of the American Society for Quality Control, staged a very fine conference. The theme of the all-day meeting was, "Quality - Its Measurement and Cost." I notice that industries refinement of most quality control techniques include cost aspects, as well as the primary objective of improving effectiveness.

This, I heartily endorse, as a taxpayer and also in my official capacity, because positive cost controls highlight waste and losses due to failure to control quality. When these truths are revealed to management, quality gets back to standard fast.

The Air Force Role

Up to this point we have dwelled upon the contractor's obligation. What then does the Air Force consider as its part in the overall quality control effort? To begin with, by public law, it is our responsibility to assure that only material conforming to contractual requirements is accepted for payment. This again is a very simplified statement of responsibility. Actually, the military part is no less complex and diverse than that outlined for industry. There is, however, a sharp distinction in the comparative extent to which the Air Force duplicates the activities of the contractor in determining acceptability of product. These are not only for economical or practical reasons, but is attributed to a very fundamental concept which became obvious a number of years ago. Very simply stated, quality and reliability must be designed and built into the product. The only way that a contractor can assure that the desired quality is built in is thru very precise step by step examination, testing and control of every important operation in the production process. To accomplish this a substantial number of people possessing a great variety of specialized talents and skills are necessary. From the Air Force point of view then, the assurance by the Government can be achieved if the effectiveness of the contractor's quality control efforts can be measured by a systematic and objective surveillance and audit of the systems, procedures and products, and the validity of quality data produced therefrom properly verified.

Why an Air Force Quality Control Specification

It was this concept which motivated the development of the Air Force Quality Control Specification MIL-Q-5923. This was necessary for two principle reasons: One, all Air Force planning, programming, training and staffing must be based on a single concept, and two, this concept would work only if specific contractual requirements for quality control were established which would be commonly applicable to all contractors. After seven years of successful operation, this decision has proven beyond any question to be a correct one. With the advent of modern weapons systems and missiles and with what we expect to see in the near future, top management must exercise absolute controls over quality and reliability and assure that these essential characteristics are built into the components comprising the weapon system. Evidence that this is accomplished by methods subject to Air Force surveillance or audit facilitates acceptance decision making and results in minimum costs for "Government inspection."

The reason then for a common effort and interest between the industry and military quality control organizations becomes obvious. It explains why the military is concerned with how the quality control manager does his job. It also explains why the military must specify, in general terms, what is required of a contractor in the way of an acceptable quality control system which assures equitable treatment

406

to all of its suppliers. It establishes the principle means by which
the military discharges its responsibility to accept materiel that is
in compliance with specified quality and reliability requirements of
the contract. It protects the citizen, both as a taxpayer and as a
competitor for Government business.

We no longer can operate on the old customer-versus-seller con-
cept. The quality control effort within the contractor's facility
becomes a matter of joint concern and cooperative relationship. It
is through these means and our participation in meetings with indus-
trial organizations, such as American Society for Quality Control,
Aircraft Industries Association, Electronic Industries Association
and similar agencies, that we can move progressively forward in the
development and improvement of these concepts and further cultivate
the military-industry team approach. Since the Air Force views these
advancements as a furtherance of its own objectives, it has and will
continue to lend full support to such efforts.

The Contractor Complies

To get back then to the question, "What does the Air Force expect
of contractors?" I must acknowledge that in many respects the expecta-
tions have already been realized. We expect, however, that you, who
are members of this great organization, which has contributed so much
to the quality of military equipment, will accept the challenge before
you now and progressively move forward in your efforts to assure even
greater effectiveness in controlling quality and reliability, which
are so vital to today's high performance aircraft, missiles, satellites
and space vehicles.

Reference:

    (1) Reliability is the probability of performing without failure
        a specified function under given conditions for a specified
        period of time (as defined by the Advisory Group on Relia-
        bility of Electronic Equipment in their report of 4 June
        1957, titled "Reliability of Electronic Equipment").

408

CONTINUOUS SAMPLING PLANS

R. E. Biedenbender
Quality Control
Headquarters, Air Materiel Command
Wright-Patterson Air Force Base, Ohio

Today I would like to review some of the activity currently under way within the Department of Defense in the field of continuous sampling. This will include a brief review of the general nature of university research sponsored by the DD under a joint service contract, a brief discussion of DD continuous sampling handbooks currently being prepared, a detailed discussion of the Air Force Multi-level Continuous Attribute Sampling Plans, and a brief review of some applications of these plans to date. Particularly with reference to the contract research mentioned above, and to internal research and development under way within the services, this report cannot of necessity represent total DD activity in the continuous sampling field. It is hoped that it will adequately serve the purpose of giving some idea of the scope and nature of the interest and work in continuous sampling today in the DD. Similarly, these remarks cover DD activity, and do not cover activity in industry and universities not participating in the joint-service contract mentioned above. I will also assume that the reader is familiar with the circumstances in which continuous sampling plans are generally most useful.

DD RESEARCH

Time and space here do not permit a detailed discussion of research recently completed or under way, and necessitate assumption that the members of the audience are familiar with the work of Dodge in continuous sampling[1]. Needless to say, of course, most of the work I will mention has generated from the early work of Dodge and Wald and Wolfowitz[2]. Research completed to date has concentrated on the following areas of continuous sampling: the effects of assuming or not assuming the process is in a state of statistical control, extension and modification of the Dodge single level procedures to include additional sampling levels, OC curve computation for continuous attribute plans employing process stop rules, processes having specified degeneration characteristics, and assessment of plans by its income characteristics, rather than the AOQ. References 3, 4, 5, 6, 7, 8, 9, and 10, at the end of this paper contain detailed discussions of this work.

Work presently under way includes investigation of the optimal level plan to choose when using multi-level attribute plans for finite production runs; examination of such multi-level plans without the assumption of the existence of the state of statistical control, and the effects of employing specific process stop rules; other various production models for continuous sampling on processes not in a state of statistical control; and the extension of continuous attribute sampling to continuous variables plans. Some work on continuous variables plans has already been done, independent of DD sponsorship[11].

DD HANDBOOKS

Turning now to a discussion of DD Handbooks concerning continuous sampling presently undergoing development, these are two. The Army Ordnance Ammunition Center is presently preparing a handbook of single-level continuous attribute sampling plans, while the Air Force is

preparing one containing multi-level continuous attribute plans. Since this latter handbook is basically an adaption of the Air Force manual (12) to be discussed in more detail during the remainder of this paper, I will confine the immediate discussion to the single-level handbook. This handbook will contain plans of two types, one following the Dodge and Dodge-Torrey, CSP-1 and CSP-2 procedures(13), currently used in the Army Ordnance Handbook(14) and the second employing a decision rule which halts further inspection until corrective action has been taken on the process to remove sources of discrepant material, as currently used in the Navy Bureau of Ordnance Handbook(15). Plans of this type are, in effect, a combination of the rejection number concept of attribute acceptance plans and the Dodge continuous plan concept. They were particularly designed to eliminate the possibility of unlimited inspection by a purchaser, which would exist if the purchaser felt it necessary to perform the 100% inspection sequences if a straight Dodge procedure were utilized. The plans are described by the parameters N, the estimated number of items in a production interval (usually a day); i, the number of consecutive items inspected which must be acceptable before sampling can begin; f, the sampling rate; and a, the maximum number of defectives allowed to occur during the production interval, before inspection is halted. Note that i and f are identical to the parameters of the Dodge plan. Inspection begins during the production interval on a 100% basis. If "a" plus one defectives occur before i successive acceptable items are observed, inspection is halted until corrective action is taken. If "a" or less defectives occur before i successive items are found, sampling at rate f is initiated. Upon occurrence of a defective, 100% inspection is reinstated. This is continued until (a) i successive acceptable items are observed, or (b) a total of "a" plus one defectives in the production interval are observed. This brief description covers the most important concept embodied in these plans. Time does not permit discussion of other important possible facets of the plans, such as reduced and tightened inspection, multi-station inspection, and action in the event of critical defects.

AMCM 74-23: BACKGROUND
Turning now to the Air Force multi-level continuous attribute manual, currently being converted to a DD handbook, Air Force interest in such plans began when a plan of this type was implemented on the green-run teardown of newly procured aircraft engines, although the theory for such a plan was not fully developed at this time. This type of plan seemed to offer possibilities of decreasing the amount of inspection or testing necessary when the quality history continues to show a process average much better than the AOQL specified; it might also serve to soften the impact of possible fluctuations between 100 percent inspection and sampling, when only one sampling level is employed.

The engine application above proved successful, and it appeared that such plans could also be useful in other expensive inspection or test areas. Consequently, when satisfactory solutions to the theory appeared in 1954(5)(16), the Air Force let a contract to Stanford University to prepare a manual, containing a complete set of multi-level plans, for use by personnel without sophisticated statistical backgrounds.

In the development of this manual, Stanford University personnel visited a number of Air Force operated installations as well as DD contractors to collect data in order to evaluate the various possible

410

statistical models for maximum simplicity of administration and operation. Actual service tests were conducted at an Air Force overhaul installation before the model was finalized.

GENERAL DESCRIPTION

AMC Manual 74-23 sampling plans provide a procedure in which the fraction inspected, of the presented items, is decreased or increased by steps or levels, dependent upon the quality of the presented items. When application to a product is started, 100% of the product is inspected until a specified number of consecutive items, i, is found free of defects. At that time, the inspection is reduced to a fraction, f, of the items presented. If i successively inspected items are found free of defects, the inspection rate is reduced to a new fraction, f^2, of the items presented. This procedure is continued through the several levels for the plan. The use of a constant value of i at all levels of a particular plan simplifies administration, and lessens the possibility of inspection error in applying the procedure. Use of a geometric progression in sampling rate, permits different sampling frequencies for different characteristics without administrative burden. It also permits use of different AOQLs for major and minor characteristics. However, if at any level a defective item is found before i successively inspected items are passed, the plan provides a "spottiness" check to determine whether sampling should continue at the present rate or revert back to the previous sampling rate.

The spottiness check is similar to the type of protection used against spottiness by the Dodge CSP-3 plans. (A very concise overall description of the plans would be an extension of Dodge's CSP-3 plans to additional sampling levels.) An examination of the schematic diagram (Figure 1) will best give the reader an understanding of the procedure followed. Although this schematic shows only three levels, plans with from one to five levels are contained in the manual. To obtain a schematic for a different number of levels, simply add or subtract to the middle levels.

The plans are described by three parameters, i, k, and f. The number of successively inspected items that must be found acceptable in order to shift to a smaller sampling rate is i. The number of sampling levels that may be used is given by k, and varies from one to five. The first sampling rate after leaving 100% is f, and each succeeding sampling rate is the same fraction raised to one larger power. Thus, for a five level plan, the inspection rates are 100%, f, f^2, f^3, f^4, and f^5. If f = 1/2, the successive rates are 100%, 1/2, 1/4, 1/8, 1/16, and 1/32. If the number of levels is less than five, the last rate corresponds to the rate at the last level, i.e., if three levels are used, the last sampling rate is 1/8. F values of one-half and one-third are planned for the DD Handbook version. The method of risk indexing employed is the traditional AOQL (Average Outgoing Quality Limit). The underlying mathematical theory can be set up employing the Markhov chain process.

RANGE OF PLANS

Figure 2 shows the plans available using f equal to one-half. It should be noted that for the same AOQL each additional level requires an increase in the value of i.

411

AFI AND AOQ CURVES

The AFI and AOQ curves for the plans reveal some interesting characteristics of the plans. Figure 3 shows the AOQ curves for the same AOQL plan with different k values. Figure 4 shows the AFI curves for the same AOQL with different k values. For either very good or very bad quality with respect to the AOQL, the highest level plans are the most efficient since they tend to either minimize or maximize inspection as the quality would logically dictate. This is, of course, consistent with general experience with lot-by-lot attribute sampling plans when considering single, double, or multiple plans.

SELECTION OF A PLAN

One of the problems historically, in continuous sampling plans, has been the logical selection of the i and f values for a plan after the AOQL has been set. In these plans, since f is already set, this problem shifts to the question of what level plan should be chosen. This can easily be done when the production quantity is exceedingly large, since the AFI curves would then be good approximations. However, it is also obvious that for small finite quantities, more inspection could be required if, for example, a k equal two plan were used instead of a k equal one plan; this could be true even though quality was very good. Availability of good quality histories reduces this problem somewhat, since it might then be feasible to begin sampling at one of the lower levels. However, since extensive records are not always available, a better solution for selecting the optimum number of levels to use for a finite production run was desired. In studying this problem, it was decided to concentrate on solutions which would minimize inspection when quality was good. Theoretical studies resulted in equations of such complexity they had to be abandoned. Monte Carlo techniques were then employed, with satisfactory results. Figure 5 shows the average fraction inspected for the different k levels for increasing sizes of production runs. The incoming quality is assumed to be equal to the AOQL.

Decision rules for level selection based on curves such as these are incorporated into a table in the manual (Figure 6). Only curves for f equal one-half and p equal to the AOQL are presently contained in the manual. It is planned to include a similar table and curves for plans with f equal to one-third in the DD Handbook version.

APPLICATIONS AND EXPERIENCE TO DATE

Since the printing of the manual in September 1956, a sizeable number of successful applications have been implemented within the Air Force. The size and nature of items included in these applications range from aircraft to EAM cards. They have also been used in controlling processes such as preservation and packaging. In the case of aircraft, of course, this is done only when a contractor has flown each individual aircraft, and has a good quality record. While it is not feasible to go into great detail on these various applications, a brief description of one application to aircraft propellers will be given. As a result of a number of field complaints of failures of one type of propeller very early in their service life, an engineering decision was made to subject all such propellers to a 10 hour reliability test simulating early service operating conditions. Conferences between engineering and quality control personnel resulted in the authorization of a multi-level sampling plan in lieu of the proposed 100% testing. This

412

resulted in a decrease of $60,000 in the first year on the initial contract in which the procedure was incorporated. The plans have since been placed in other propeller contracts.

The plans have proven quite useful in Air Force Depot maintenance overhaul and supply storage operations, both from a process as well as a product control viewpoint, since the nature of the plant layouts and the products overhauled are better suited for continuous, rather than lot-by-lot, sampling. The number of applications to relatively simple items has been surprising since it was originally believed that the administrative complexity would tend to result in concentration of applications in expensive test areas on the more complex type of product.

A slight juggling of the plans' mathematical model has also resulted in a model for Defects-per-Unit continuous plans. Such a plan can easily be set up along traditional continuous attribute lines by defining an unacceptable unit to be one with more than some arbitrary number of defects, and assuming that defects follow the Poisson distribution(17, 18). These are currently undergoing service testing, and are thus not yet in finalized form. Such plans, just as the traditional C-chart, offer considerable flexibility in defining the unit of product, resulting in simplicity of administration and operation. One unexpected application to date, making use of this feature, is EAM card verification. The unit of product is defined to be 100 successive cards. Service test data verified that the error rate of better operators follows the Poisson distribution assumption.

In summary, let me repeat that this discussion is not intended to cover all the activity within the DD in the field of continuous sampling. It is hoped, however, that this discussion has given a broad picture of the scope and nature of the interest in this field.

References

1. H. F. Dodge, "A Sampling Inspection Plan for Continuous Production," Am. Math. Stat., Vol 14 (1943) pp 264-279.
2. A. Wald and J. Wolfowitz, "Sampling Inspection Plans for Continuous Production Which Insure a Prescribed Limit on the Outgoing Quality" Am. Math. Stat., Vol 16 (1945) pp 30-49.
3. Gerald J. Lieberman, "A Note on Dodge's Continuous Inspection Plan," Tech. Rep. No. 12, Applied Mathematics and Statistics Laboratory, Stanford University, 1953.
4. M. A. Girshick, "A Sequential Inspection Plan for Quality Control", Tech. Rep. No. 16, Applied Mathematics and Statistics Laboratory, Stanford University, 1954.
5. Gerald J. Lieberman and Herbert Solomon, "Multi-Level Continuous Sampling Plans," Tech. Rep. No. 17, Applied Mathematics and Statistics Laboratory, Stanford University, 1954.
6. I. Richard Savage, "A Three Decision Continuous Sampling Plan for Attributes," Tech. Rep. No. 20, Applied Mathematics and Statistics Laboratory, Stanford University, 1955.
7. C. Derman, S. Littauer, H. Solomon, "Tightened Multi-Level Continuous Sampling Plans," Tech. Rep. No. 28, Applied Mathematics and Statistics Laboratory, Stanford University, 1956.

8.   I. Richard Savage, "Statistical Production Models and Continuous Sampling Plans," Tech. Rep. No. 29, Applied Mathematics and Statistics Laboratory, Stanford University, 1956.

9.   Geoffrey Gregory, "An Economic Approach to the Choice of Continuous Sampling Plans," Tech. Rep. No. 30, Applied Mathematics and Statistics Laboratory, Stanford University, 1956.

10. Craig Magwire, "Finite Continuous Sampling Plans Involving a Stopping Rule," Tech. Rep. No. 31, Applied Mathematics and Statistics Laboratory, Stanford University, 1956.

11. Harry G. Romig, "New Statistical Approaches in Aviation and Allied Fields," Proceedings, Aircraft Quality Control Conference, 1953.

12. AMCM 74-23, "Multi-Level Continuous Sampling Acceptance Plans for Attributes."

13. H. Dodge and M. Torrey, "Additional Continuous Sampling Plans," Industrial Quality Control, VII, No. 5, 1957, pp 7-12.

14. ORD-M608-11, "Procedures and Tables for Continuous Sampling by Attributes," Ordnance Inspection Handbook.

15. NAVORD STD-81, "Sampling Procedures and Tables - For Inspection by Attributes on a Moving Line (Continuous Sampling Plans)."

16. Joseph A. Greenwood, "A Continuous Sampling Plan and its Operating Characteristics," Bureau of Ordnance, Navy Department, unpublished memorandum.

17. John E. Condon, "Continuous Defects-per-Unit Sampling Plans," Headquarters, Air Materiel Command, USAF, unpublished memorandum.

18. Richard E. Biedenbender, "Multi-Level Continuous Defects-Per-Unit Sampling Plans," Headquarters, Air Materiel Command, USAF, unpublished memorandum.

FIGURE 1 – FLOW PROCESS CHART FOR MULTI-LEVEL
CONTINUOUS SAMPLING PLANS
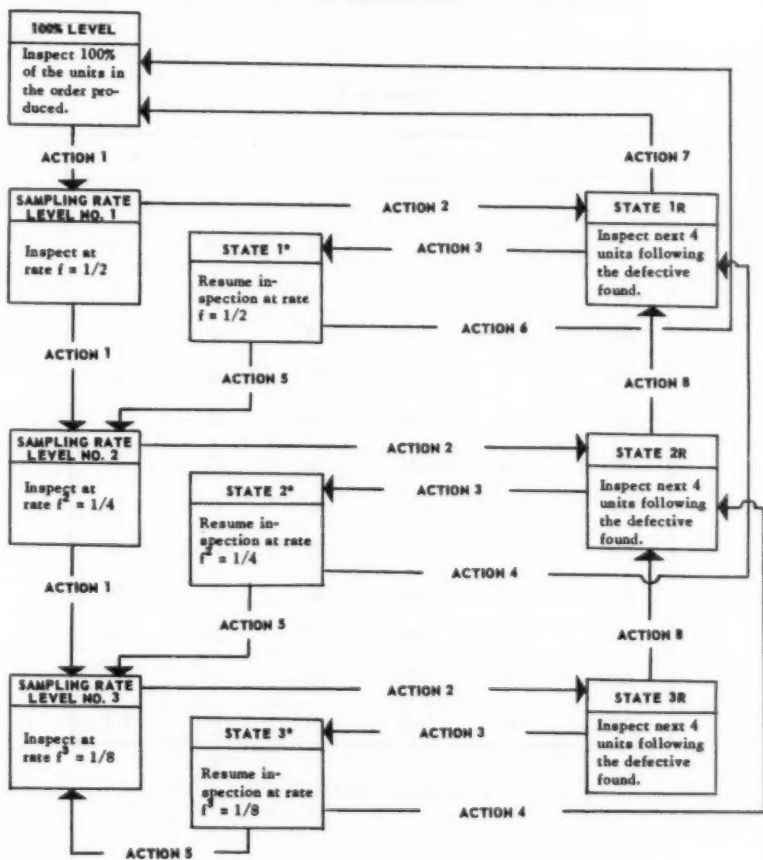
ACTION 1 – If i inspected units are found
to be free of defectives, shift
to next level.

ACTION 2 – If a defective is found while
sampling at rate f^K, shift to
State KR.

ACTION 3 – If the 4 inspected units are
found to be free of defectives,
shift to State K.°

ACTION 4 – If a defective is found after
resuming inspection at rate
f^K, shift to State (K-1)R.

ACTION 5 – If i-4 inspected units are
found to be free of defectives,
shift to next level.

ACTION 6 – If a defective is found after
resuming inspection at rate f =
1/2; shift to 100% inspection.

ACTION 7 – If a defective is found among
the 4 inspected units, shift
to 100% inspection.

ACTION 8 – If a defective is found among
the 4 inspected units, shift
to State R.

415

FIGURE 2 – TABLE OF I VALUES FOR MULTI-LEVEL CONTINUOUS ATTRIBUTE PLANS

Based on $f = \frac{1}{2}$

AOQL	k = 1	k = 2	k = 3	k = 4	k = 5
.20	–	–	–	–	–
.15	–	–	–	4	5
.10	–	4	6	8	9
.075	–	6	9	11	13
.050	5	11	15	18	20
.040	7	14	19	22	25
.030	11	20	26	31	34
.020	18	31	40	47	51
.015	25	43	55	63	69
.010	39	65	83	95	104
.0075	54	88	112	128	140
.0050	82	133	168	193	210
.0025	167	269	337	386	422

k = number of sampling levels.

416

FIGURE 3 - AVERAGE OUTGOING QUALITY CURVES
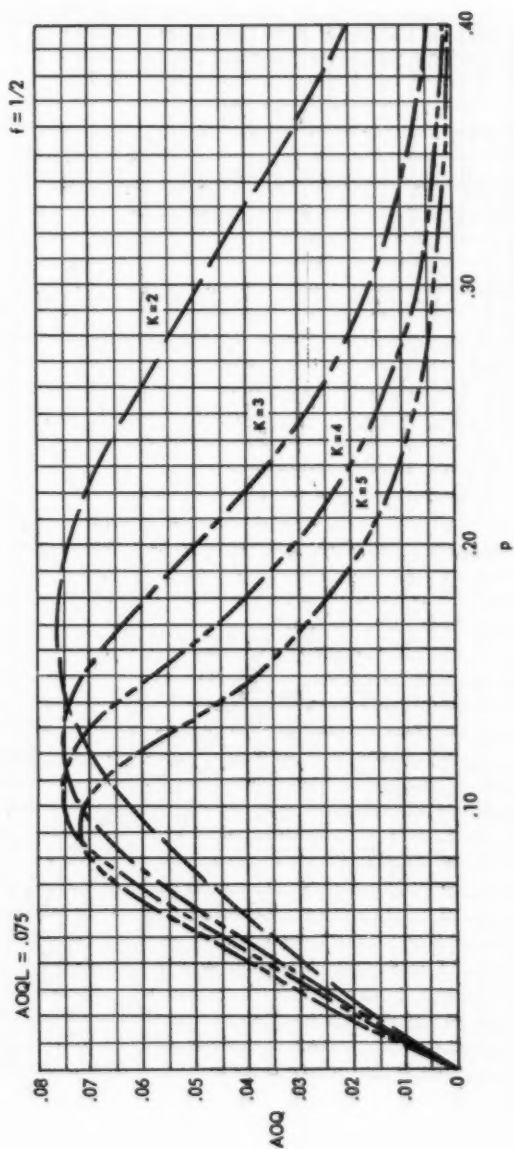
FIGURE 4 - AVERAGE FRACTION INSPECTED CURVES
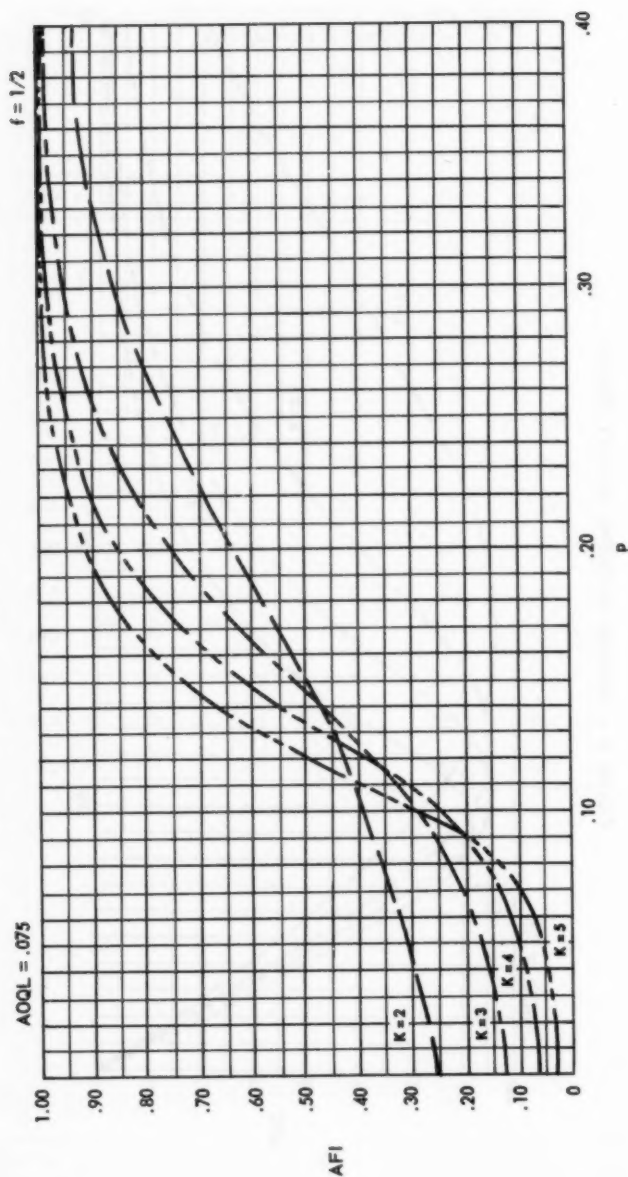
FIGURE 5 - ESTIMATE OF AVERAGE FRACTION INSPECTED FOR
FINITE LOT SIZES, WHEN INCOMING QUALITY IS
EQUAL TO THE AVERAGE OUTGOING QUALITY LIMIT
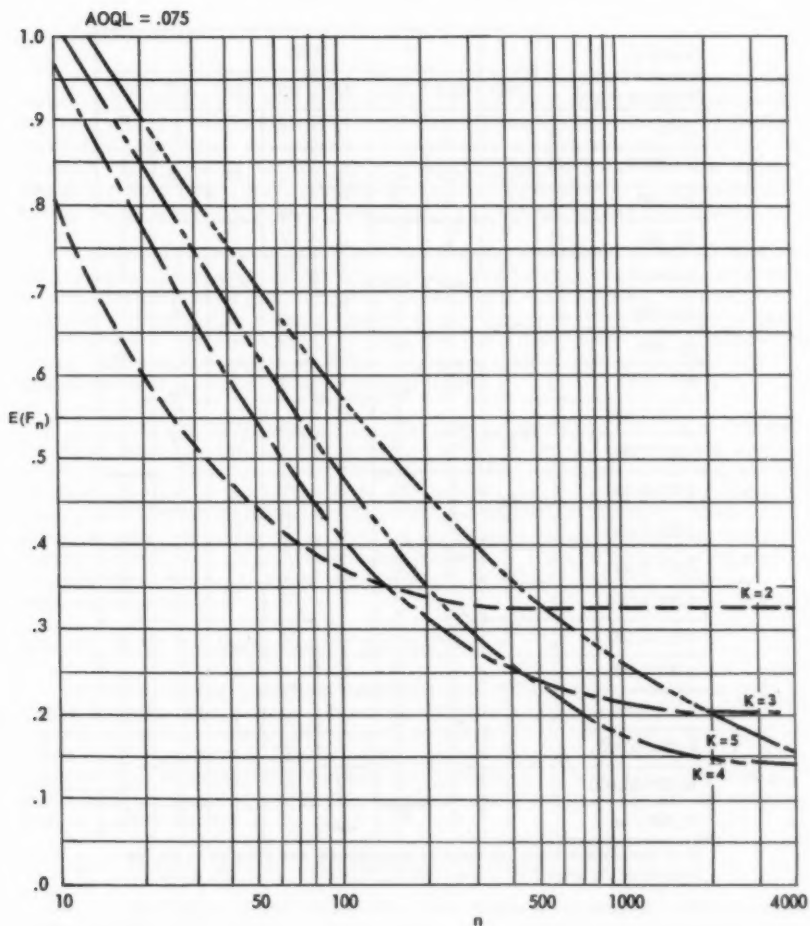
FIGURE 6 – MULTI-LEVEL CONTINUOUS ATTRIBUTE PLANS

Based on $f = \frac{1}{2}$

Contract Size or Annual Production Rate	i/k	AOQL .010	.015	.020	.030	.040	.050	.075	.100
25-50	i/k	39/1	25/1	18/1	11/1	7/1	11/2	6/2	4/2
51-100	i/k					14/2			6/3
101-200	i/k				20/2			9/3	
201-300	i/k		43/2	31/2		15/3			
301-400	i/k	65/2				19/3			
401-600	i/k				26/3				8/4
601-800	i/k			40/3				11/4	
801-1,000	i/k		55/3			22/4	18/4		
1,001-1,400	i/k				31/4				
1,401-1,800	i/k	83/3							
1,801-2,200	i/k			47/4					9/5
2,201-2,600	i/k							13/5	
2,601-3,000	i/k			63/4					
3,001-4,000	i/k						20/5		
4,001-5,000	i/k	95/4				25/5			
5,001-6,000	i/k								
6,001-8,000	i/k								
8,001-10,000	i/k								
10,001-15,000	i/k				34/5				
15,001 - up	i/k			51/5					

If on entry into a block, an arrow is encountered, use the plan at the top or point of the arrow.

EVALUATION OF TUBE SCREENING PROCEDURES

PART I - ENGINEERING EVALUATION

Vinson M. Lockwood
Hughes Aircraft Company*

1.0 INTRODUCTION

Faced with ever-increasing demands for more reliable equipment, many contractors engaged in the manufacture of military electronic equipment, particularly in the aviation and missile fields, have resorted to parts screening techniques as a means of enhancing the reliability of their products. To a large extent, this screening may be traced to the presumption that future part performance in the equipment will be enhanced if the parts have previously conformed to these screening procedures.

Insofar as electron tubes are concerned, considerable effort has been devoted to research, development and production refinement in order to improve the basic quality level of military tube types. However, equipment contractors paid little heed to the great strides made in the improvement of tube reliability over the past five years and continued to employ their rigid incoming inspection test procedures to screen individual defective tubes from purchased lots.

As a result of this practice, incoming inspection shrinkage figures ran high, and the cost of "acceptable" tubes mounted. When asked for substantiating data to justify these practices, no individual contractor could supply any. Thus, the need for an evaluation of these screening procedures became apparent in order to justify the additional cost of electron tube procurement as well as to determine whether or not they should be included in military specifications. The military services, through the Bureau of Ships, requested that ARINC undertake such an evaluation program.

2.0 SCOPE OF THE ARINC EVALUATION PROGRAM

Three specific incoming inspection procedures were chosen as subjects for this evaluation. These particular procedures were representative of tube screening procedures in general, in that they included tests involving visual (microscopic) inspection, X-ray examination, polariscopic examination, thermal shock testing, electrical inspection, and vibration testing. The three procedures will hereinafter be referred to as Procedures A, B, and C.

2.1 Description of the Screening Procedures

Procedure A consisted of a two-stage microscopic inspection and a swept frequency vibration noise test. The microscopic inspection criteria were derived from BuOrd Specification MIL-E-17751; the vibration test was composed of peak-to-peak noise measurements under swept frequency excitation from 30 cycles per second to 1000 cycles per second at an acceleration of 15 g's.

Procedure B was primarily a thermal shock screening test. This test was preceded by a visual inspection for glass defects, as described in MIL-E-17751. Then the tubes were subjected to three cycles of the MIL-E-1 thermal shock test. For those not familiar with the MIL-E-1 "dunk" test, it consists of placing room temperature tubes into boiling water (not lower than 97° C) for a minimum of fifteen seconds and then immediately transferring them to ice water (not higher than 5° C) for five seconds. This procedure included three such "dunkings", with a minimum of fifteen seconds between each "dunk". Forty-eight hours after this thermal cycling had been completed, the tubes were again inspected for glass defects in accordance with MIL-E-17751.

Procedure C included visual inspection, X-ray examination, polariscopic examination, vibration testing, and electrical testing. Visual inspection was performed essentially in accordance with MIL-E-17751, and vibration and electrical tests were performed in accordance with the requirements of MIL-E-1 applicable to the individual tube types. In addition to these standard tests, X-ray examination was employed to distinguish between metallic and non-metallic particles and to detect defective welds, and polariscopic examination was employed to detect excessive strain within the straight portion of the glass side walls.

2.2 Description of Test Designs

In evaluating each of these procedures, experiments were designed which had as their primary objective the comparison of screened tubes with unscreened tubes. Also included were studies pertaining to test repeatability, tube deterioration, correlation of test criteria with equipment performance requirements, NEL tube shield inserts, and improved tube types. Only those specific objectives pertaining to the subject at hand will be further pursued at this time.

The sample sizes employed for evaluating each procedure were based upon performance estimates which had been observed beforehand as well as test sensitivity and test power. Thus, in obtaining answers to the primary objective, the performance of two samples was monitored. The treated sample was composed of tubes which had passed the screening procedure. The control sample consisted of tubes selected at random beforehand from the same lots but not subjected to the screening procedure; and in order to prevent any extraneous factors from influencing the test results, the principle of randomization was utilized in each evaluation. In addition, sample identity was unknown to the maintenance and test personnel.

2.3 Description of Equipment Environments and Performance Measurements

For evaluating Procedure A, tubes from each sample were installed and operated in a missile component which was subjected to electrical and vibration tests conducted in accordance with the component production test specification. Those tubes causing component non-compliance to test specifications, as evidenced by their removal from an assigned socket by production personnel, were termed "removals". Based upon the total number of tubes installed within each sample, removal percentages were calculated.

For evaluating Procedure B, tubes from each sample were subjected to simulated thermal environments which were representative of the
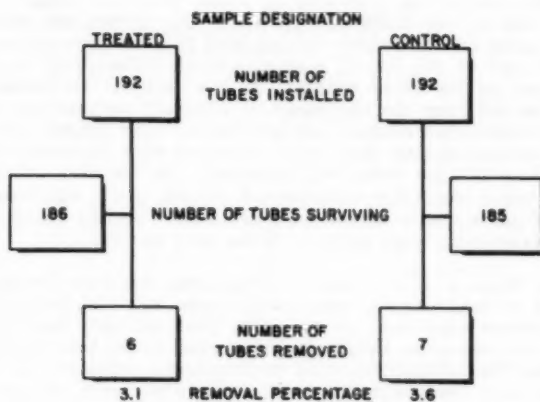
Fig. 1   Evaluation of Procedure A
(Missile Production Testing)

O – TREATED SAMPLE
X – CONTROL SAMPLE
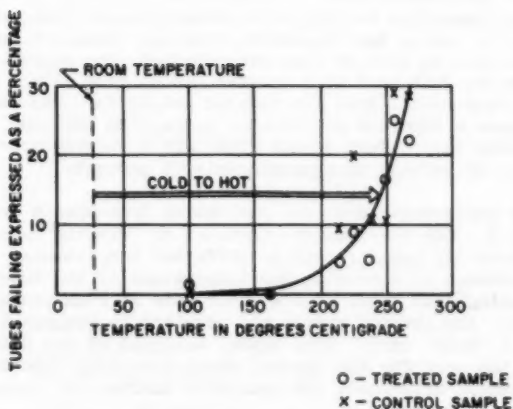
Fig. 2   Evaluation of Procedure B
(Cold-to-Hot Thermal Shock)

423

thermal environments encountered in production "potting" of electrical units. After a 48-hour holding period, electrical tests for plate current and grid current were employed to determine "removals". Here, as before, based upon the total number of tubes tested within each sample, removal percentages were derived.

For evaluating Procedure C, tubes from each sample were installed in each of two different systems. One system was made up of electronic units of the BomNav system used in B-47 type aircraft. The transceiver unit of the ARC-27 communications system used in AD type aircraft was used as the other system. In the case of the former system, all tubes removed from the equipment by military technicians in accordance with established maintenance procedures were termed failures. Based upon the assumption that these failures were distributed exponentially,[1] tube failure rates were derived. In the case of the latter system, all units requiring maintenance action, i.e., adjustment or replacement of parts, were termed unit failures. Again, based upon the exponential assumption, unit failure rates were derived.

Of these four evaluative techniques, two were studies of initial circuit or environment capability, since the time periods during which performance measurements were made included only the time necessary to ensure component compliance with production test specifications. The other two evaluations included deterioration studies, in that time on each tube or unit was monitored until failure occurred, as evidenced by tube replacement or unit maintenance action.

3.0 RESULTS OF TESTS

The results obtained in the ARINC evaluation program reveal that no apparent improvement in tube performance or equipment performance is accomplished through the use of these tube screening procedures.

The tube performance data for evaluating Procedure A which were obtained during production testing of a guided missile component is presented in Figure 1. Using the sequential sampling technique, which is covered in the second part of this presentation, the statement that no difference exists that is 2 to 1 greater can be made with a Type II error of 0.05. You may recall that the treated sample for this evaluation was required to pass a rigorous microscopic inspection test and a swept frequency vibration test. Tube yields from this screening procedure ranged from as low as 10 percent to approximately 75 percent.

The tube performance data for evaluating Procedure B is presented in Figures 2 and 3. One environment consisted of "dunking" tubes at room temperature into oil baths raised to different temperatures (cold-to-hot); the other consisted of elevating the temperature of the tubes in an oven and then "dunking" the tubes into a water bath at room temperature. Here, you may recall, the treated sample was required to withstand three cycles of the MIL-E-1 "dunk" test. Here again, analysis of the data fails to substantiate the need for the thermal shock screening procedure in that the control sample withstands the simulated thermal environments just as well as the treated sample.

Figure 4 presents the tube performance data for evaluating Procedure C when measured in the environments associated with the BomNav system installed in B-47 aircraft. In this case, you should bear in mind that the treated sample was required to successfully meet the requirements imposed

424
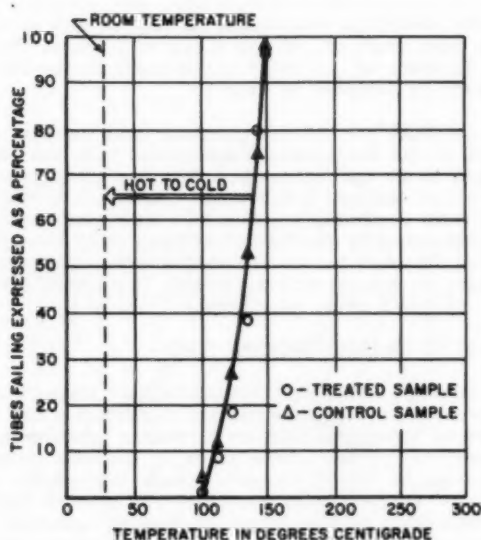
Fig. 3   Evaluation of Procedure B
(Hot-to-Cold Thermal Shock)

	SAMPLE DESIGNATION	
	TREATED	CONTROL
TOTAL TUBE HOURS	1,749,343	1,676,710
NUMBER OF REMOVALS	126	117
FAILURE RATE (PER CENT PER 100 HOURS)	.72	.70

Fig. 4   Evaluation of Procedure C
(BomNav System)

	SAMPLE DESIGNATION	
	TREATED	CONTROL
TOTAL UNIT HOURS	12,333	11,448
NUMBER OF UNIT MAINTENANCE ACTIONS	227	219
MEAN TIME-BETWEEN UNIT MAINTENANCE ACTIONS	54.3	52.3

Fig. 5   Evaluation of Procedure C
(Communications System)

by visual, X-ray, polariscopic, electrical, and vibration testing; and as evidenced in the preceding evaluation, no improvement in tube performance was noted at a test power of .95 for a 1.3 to 1 ratio. For a 1.1 to 1 ratio, a Type II error of .15 would be involved if the statement that no difference exists is accepted as true.

Figure 5 presents the unit performance data for evaluating Procedure C when measured in the environments associated with the communications system installed in AD type aircraft. This was another evaluation of Procedure C, but it employed a different "yardstick", so to speak. The effects of the tube screening procedure on tube performance were paramount in the evaluation of the BomNav system, while the effects on unit maintenance were foremost in the evaluation using the communication system. Here again, no improvement was noted. This statement involves a test power of .95 for a ratio of 1.3 to 1.

4.0 CONCLUSIONS TO BE DRAWN FROM THE TESTS

In brief summary, the four evaluations have demonstrated that no material improvement in tube or equipment performance has been attained through the use of these special tube screening procedures, in that the so-called "defective" product seems to enjoy the same probability of survival as the so-called "good" product. Since the results are contrary to the predictions of the screening procedure, it appears logical that other factors must contribute to the degree of tube or equipment reliability inherent in any system. This will become more apparent after we ponder over the following considerations:

4.1 Test Criteria

For the most part, screening procedures judge the "goodness" of a product by criteria that are not necessarily related to the criteria by which a system judges the "goodness" of a product. For anyone with even the remotest experience in electronics, the following conversation between the two individuals depicted in Figure 6 would appear commonplace:

Bill: "But I tested this tube myself, and I know it's good!"

Sad Sack: "I don't care who in the H--- tested the blankety-blank tube! It won't work!"

4.2 Repeatability

In some cases, repeatability of a test could allow "defective" product to be present in the so-called "good" product. An additional objective included in the evaluation of Procedure A was to determine the repeatability of the visual inspection test. It was discovered that the agreement between the tests on individual tubes in a sample was of the order of 70 per cent, whereas the agreement between the tests on percent defective in the sample was of the order of 90 percent. Here the same operator performed the two tests; if the operator-to-operator repeatability is considered, agreement between tests on individual tubes in a sample would be even further reduced. If a particular test criterion is correlated with equipment performance requirements, this test repeatability on an individual tube classification should probably exceed 90 per cent for it to be acceptable as a screening test.

426

Fig. 6　　　" . . . it won't work!"

4.3 Test deterioration

Another consideration that should not be passed over lightly is the possibility that the screening procedure may introduce deterioration which cancels out the improvement gained by the screening tests. Thus, useful life may be enhanced by the test criteria but, at the same time, be degraded by the testing and handling involved in the procedure.

4.4 Sampling

There is an old adage - "Quality must be built in . . . it can not be tested into a product." If the basic quality level of a production lot of tubes falls below the minimum acceptable level, it is unlikely that any amount of special processing by the tube user can screen out the defectives from the lot and thus test the desired quality level into the lot. The very tests which have proven ineffective for screening purposes have been found effective in assessing lot quality level and thereby indicating to a certain extent the degree of reliability to be expected in any system. Tests of this type are not only important to the tube manufacturer for good quality control, but are equally useful to the tube user in lot acceptance on a sampling basis. Some equipment manufacturers actually maintain a Vendor Rating System on the basis of such sampling plans. Economic pressure can then induce the tube manufacturer to better the quality level of his product beyond specification minimums.

427
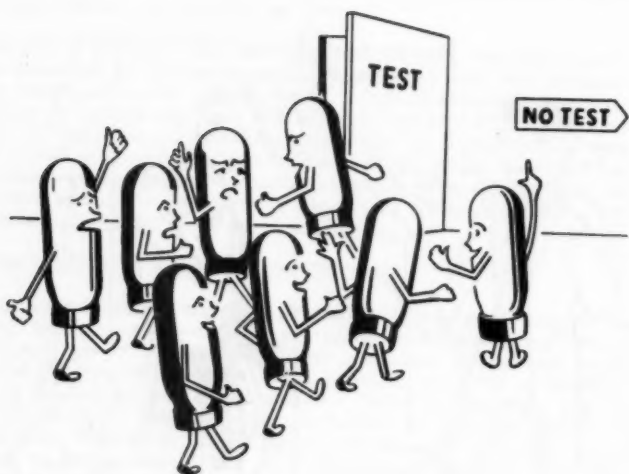
Fig. 7
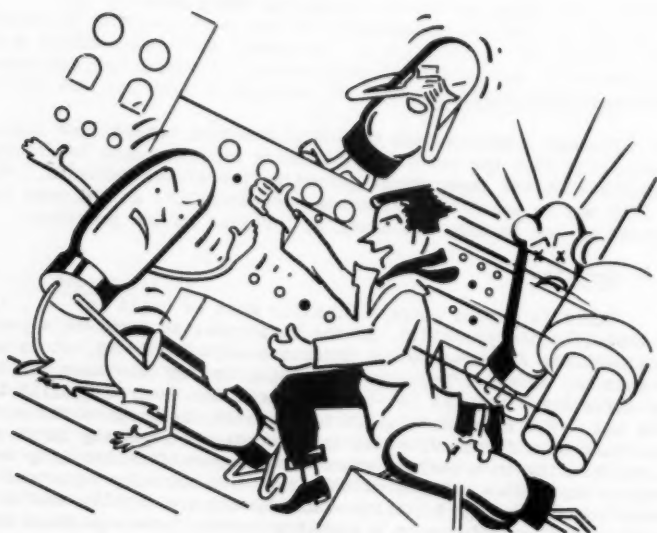
Fig. 8

428
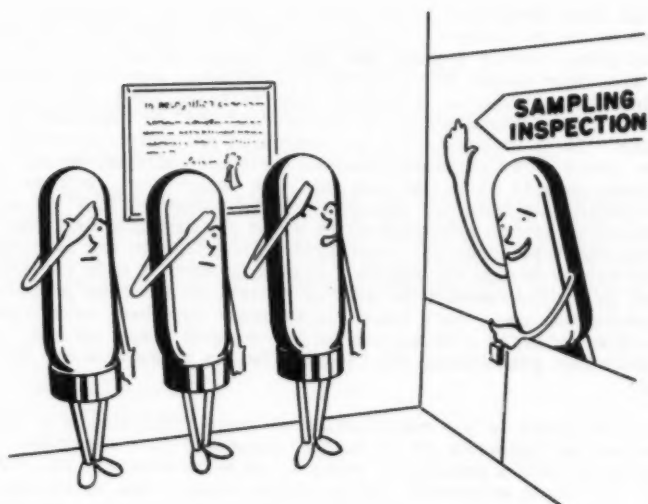
Fig. 9
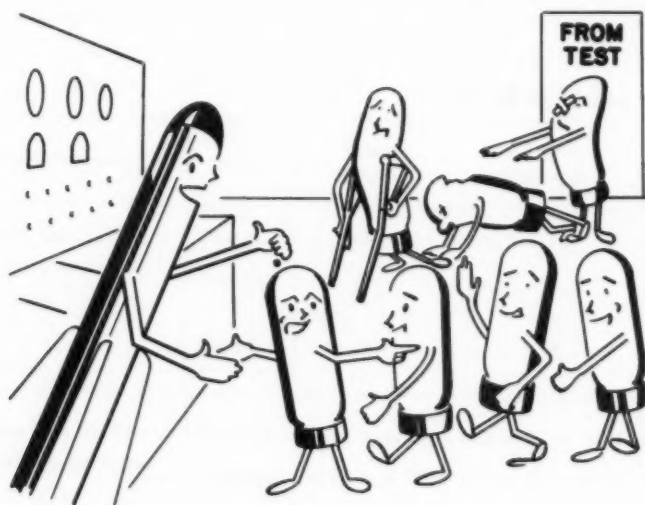
Fig. 10

5.0 A SUGGESTED APPROACH TO FUTURE SCREENING PROGRAMS

In the final analysis, it has been the consistent experience of many investigators in the field of reliability that tolerance problems generated by marginal circuit designs and basic misapplications of tubes far outweigh all other causes of equipment malfunction. For every equipment malfunction attributable to "inherent" tube defects, there are usually three to ten malfunctions which are either "induced" or "defined" by the application. It is doubtful that any screening procedure, no matter how elaborate, could offer a general, across-the-board solution to all tolerance problems as well as to the inherent tube weaknesses. Infofar as future screening programs are concerned, it is suggested that the tolerances imposed by system requirements be given primary consideration. Each circuit must be carefully considered in relation to the requirements imposed by system design tolerances. Allowance must be made for initial variations in part parameters as well as variations in these parameters during operational use. Once these requirements have been established, they should be verified - first, at the bread-board stage; second, during limited prototype production; and finally, during production and operational use.

If at any point in the verification of these requirements a screening procedure is indicated, it is further suggested that coincident with the adoption of such a procedure, steps be taken to evaluate its effectiveness. If proven effective, the procedure should then be incorporated into standard part specifications. If ineffective, it should be discontinued to conserve material and manpower. With this type of approach, based on sound engineering judgement and supported by invaluable statistical evidence, the path to a more reliable future should be brightened, indeed!

(1) ARINC Monograph No. 1

PART II. STATISTICAL ANALYSES OF DATA

Edward Sax
Aeronautical Radio, Inc.

1.0 ANALYSIS OF PROCEDURE A (MICROSCOPIC EXAMINATION, NOISE TEST)

1.1 Introduction to the Sequential Chi-Square Test

Procedure A was evaluated by comparing the proportion of removals among the tubes accepted under Procedure A (P_A) with the proportion of removals among tubes in the control sample not subjected to Procedure A (P_Q). This can be stated in the form of the statistical null hypothesis (H_0) and alternative hypotheses (H_a), as follows:

$$H_O: \quad P_A = P_Q = P_O \quad \text{(any percent)} \quad (1)$$

against the two-sided alternative hypotheses:

$$H_a: \quad P_A = P_0 + \delta/2, \quad P_Q = P_0 - \delta/2$$

$$\text{or} \quad P_A - P_Q = \delta \quad (2)$$

where δ can be positive or negative. The expected frequencies for both hypotheses are given in their respective contingency tables, in Table 1. (The "control" group is the group of tubes not subjected to Procedure A, while the "treated" group is the group accepted under Procedure A.)

TABLE 1. CONTINGENCY TABLES: EXPECTED FREQUENCIES				
Expected Frequencies	Group	Tubes Removed	Tubes Retained	Total
Under the null hypothesis (M_1)	Control	nP_0	$n(1 - P_0)$	n
	Treated	nP_0	$n(1 - P_0)$	n
	Total	$2nP_0$	$2n(1 - P_0)$	2n
Under the alternative hypotheses (M_1')	Control	$n(P_0 - \delta/2)$	$n(1 - P_0 + \delta/2)$	n
	Treated	$n(P_0 + \delta/2)$	$n(1 - P_0 - \delta/2)$	n
	Total	$2nP_0$	$2n(1 - P_0)$	2n

From the contingency tables is derived the non-centrality parameter, λ, of the non-central χ^2 distribution:[1]

$$\lambda = \sum \frac{(M_i' - M_i)^2}{M_i} = \frac{2\left(\frac{n\delta}{2}\right)^2}{nP_o} + \frac{2\left(\frac{n\delta}{2}\right)^2}{n(1-P_o)} = \frac{2n\delta^2}{4P_o(1-P_o)} \qquad (3)$$

When $\delta = 0$, the null hypothesis is true, and the non-central χ^2 becomes the χ^2.

To test the null hypothesis, the removal percentages P_A and P_O are compared by the chi-square test. If the observed χ^2 exceeds its critical value with a prescribed probability (α, for example), then the null hypothesis is rejected -- but with the understanding that a Type I error of size α could have occurred, since, when the null hypothesis is true, critical values of χ^2 will occur α percent of the time. On the other hand, if the observed χ^2 does not exceed its critical value, it may be concluded either that no difference exists between the two populations, or that an existing difference was not detected, and a Type II error of size β occurred. β is the probability of failing to reject the null hypothesis when the alternative hypothesis is true.

Since the experimenter faces the probability of making either type of error, he must decide, before the experiment, the size error of each type that he is willing to tolerate. When this is done, the sample size can be determined, to give the "power of the test" -- which is defined as the probability of detecting a difference (δ) for a given Type I error. Its relation to the Type II error is:

$$\text{Power} = 1 - \text{Type II Error} = 1 - \beta. \qquad (4)$$

The sample size for each group is calculated from Equation (3), when δ and P_o are given in the alternative hypotheses, and λ is that parameter of the non-central χ^2 that will ensure the desired power. The values of λ are computed in Tables of Noncentral χ^2, by Evelyn Fix.[2] Since the required sample size may be large, a sequential test procedure may be beneficial. The sequential test is designed so that, for a given α, the null hypothesis will be accepted if $\beta(n)$ is less than, or equal to, $\beta(N)$ -- where n is the sample size of the sequential test, and N is the sample size of the non-sequential test that gives the desired power. If $\beta(n)$ is greater than $\beta(N)$ for any value of n, additional observations are taken. This procedure is presented graphically in Figure 1. When the null hypothesis is true, the χ^2 value will fall in the "region for rejecting null hypothesis," α percent of the time; when it is false, the value will fall in the "acceptance region," β percent of the time. If the χ^2 value falls in the "region of indecision" (as it will, more than α percent of the time under the null hypothesis, and more than β percent of the time under the alternative hypothesis), additional observations are taken, and a χ^2 test is made on the pooled data.
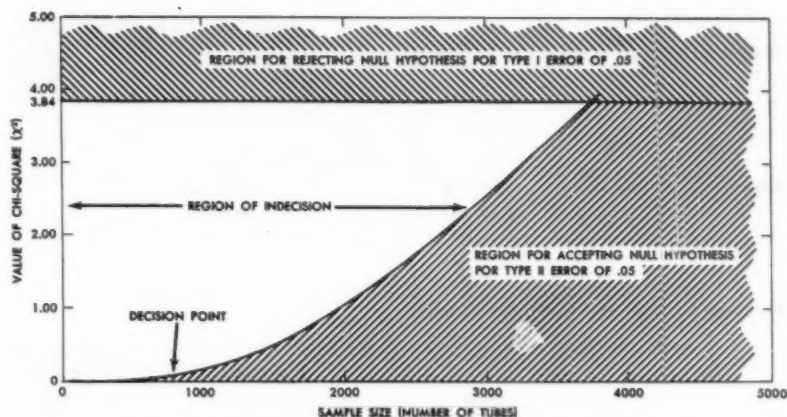
FIGURE 1. DECISION REGIONS FOR SEQUENTIAL CHI-SQUARE TEST

1.2 Method of Determining the Critical Regions

The critical region for rejecting the null hypothesis is determined by α and the degrees of freedom of χ^2. In this experiment, $\alpha = .05$, and χ^2 has one degree of freedom. Therefore, the critical region is χ^2 greater than 3.84. The critical region for accepting the null hypothesis is determined by λ_n -- found in Equation (3) -- and the critical χ^2 -- for example, χ^2_c -- that satisfied the equation.

$$\beta(n) = \int_0^{\chi^2_c} f(\dot{\chi}^2, \lambda_n)d\dot{\chi}^2 = .05 \tag{5}$$

The probability density is that of the non-central χ^2 of one degree of freedom, with non-centrality parameter λ_n. From its definition,[3/] $f(\dot{\chi}^2, \lambda)$ is derived easily.

$$\left. \begin{array}{l} \dot{\chi}^2 = X^2 \text{ where } f(X) = \dfrac{1}{\sqrt{2\pi}} \ e^{-(X-\mu)^2/2}, \ -\infty < X < \infty \\[3mm] X = \pm\sqrt{\dot{\chi}^2} \\[3mm] dX = \dfrac{d\dot{\chi}^2}{2\sqrt{\dot{\chi}^2}} \end{array} \right\} \tag{6}$$

Corresponding to a given χ^2, there are two values of X -- that is, $\pm$ X.

433

The probability element of $\dot{\chi}^2$ at $+X$ is: $\dfrac{1}{\sqrt{2\pi}} e^{-\left(\sqrt{\dot{\chi}^2}-\mu\right)^2/2} \dfrac{d\dot{\chi}^2}{2\sqrt{\dot{\chi}^2}}$

and of $\dot{\chi}^2$ at $-X$ is: $\dfrac{1}{\sqrt{2\pi}} e^{-\left(-\sqrt{\dot{\chi}^2}-\mu\right)^2/2} \dfrac{d\dot{\chi}^2}{2\sqrt{\dot{\chi}^2}}$

The sum of the two elements gives

$$
\left.
\begin{aligned}
f(\dot{\chi}^2, \lambda_n)d\dot{\chi}^2 &= \frac{1}{\sqrt{2\pi}} e^{-1/2\left(\sqrt{\dot{\chi}^2}+\mu\right)^2} \frac{d\chi^2}{2\sqrt{\dot{\chi}^2}} \\
&+ \frac{1}{\sqrt{2\pi}} e^{-1/2\left(\sqrt{\dot{\chi}^2}-\mu\right)^2} \frac{d\dot{\chi}^2}{2\sqrt{\dot{\chi}^2}}
\end{aligned}
\right\} \quad (7)
$$

Therefore, Equation (5) becomes

$$
\left.
\begin{aligned}
\beta(n) &= \int_0^{\chi_c^2} \frac{1}{\sqrt{2\pi}} e^{-1/2\left(\sqrt{\dot{\chi}^2}+\mu\right)^2} \frac{d\dot{\chi}^2}{\sqrt{2}\,\dot{\chi}^2} \\
&+ \int_0^{\chi_c^2} \frac{1}{\sqrt{2\pi}} e^{-1/2\left(\sqrt{\dot{\chi}^2}-\mu\right)^2} \frac{d\dot{\chi}^2}{2\sqrt{\dot{\chi}^2}}
\end{aligned}
\right\} \quad (8)
$$

Equation (8) is not in a convenient form for computations. In order to utilize the extensive Tables of Normal Probability Functions, [4] published by the National Bureau of Standards, the necessary transformations were made.

$$
t = \sqrt{\chi_c^2} + \mu \quad \text{and} \quad v = \sqrt{\chi_c^2} - \mu . \tag{9}
$$

Then, Equation (8) becomes

$$\beta(n) = \int_{\mu}^{\chi_c+\mu} \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \, dt + \int_{-\mu}^{\chi_c-\mu} \frac{1}{\sqrt{2\pi}} e^{-v^2/2} \, dv$$
$$= F(\chi_c+\mu) - F(\mu) + F(\chi_c-\mu) - F(-\mu)$$
$$= F(\chi_c+\mu) + F(\chi_c-\mu) - 1$$
(10)

Since $F(t)$ is the normal cumulative distribution function, $F(t)$ and $F(-t)$ are related in this manner:

$$F(t) + F(-t) = 1 \tag{11}$$

Therefore,

$$\frac{F(\chi_c+\mu) + F[-(\chi_c+\mu)]}{2} = \frac{1}{2} \quad \text{and} \quad \frac{F(\chi_c-\mu) + F[-(\chi_c-\mu)]}{2} = \frac{1}{2}$$

so that

$$\beta(n) = \frac{1}{2}\left\{ F(\chi_c+\mu) - F[-(\chi_c+\mu)] + F(\chi_c-\mu) - F[-(\chi_c-\mu)] \right\}$$
$$= \frac{1}{2} \int_{-(\chi_c+\mu)}^{(\chi_c+\mu)} \frac{e^{-t^2/2}}{\sqrt{2\pi}} \, dt + \frac{1}{2} \int_{-(\chi_c-\mu)}^{(\chi_c-\mu)} \frac{e^{-v^2/2}}{\sqrt{2\pi}} \, dv$$
(12)

From Equation (9),

$$\chi_c = \frac{t+v}{2} \quad \text{and} \quad \mu = \frac{t-v}{2} = \sqrt{\lambda} \tag{13}$$

435

From Equation (3), $\mu = \sqrt{\lambda}$ is computed; then, χ_c is computed from Equation (13) for the limits of t and v that satisfy Equation (12). If v is negative, the limits of the integral should be reversed. The value of the integral becomes negative.

1.3 Critical Regions of the Experiment

In the experiment described in Part I, the statistical hypotheses (with $\alpha = \beta = .05$) are:

$$H_o: \quad P_Q = P_A = P_o \quad \text{(any percent)}$$

$$H_a: \quad (P_Q = 4\%, \ P_A = 2\%) \quad \text{or} \quad (P_Q = 2\%, \ P_A = 4\%).$$

The alternative hypothesis implies that $P_o = 3\%$ and $\delta = 2\%$. From Equation (3),

$$2n = \frac{4\lambda P_o(1-P_o)}{\delta^2} = 3,779$$

is the non-sequential sample size, based on $\lambda = 12.985$, the parameter value for a power of 0.95. To obtain this size sample, 80 package runs were necessary; therefore, the sample size was increased to 3,840, to include all tubes from the 80 runs.

The sequential tests were made every two package runs. The tubes in both groups were assigned to sockets at random, with balance maintained for each pair of test runs and, insofar as practical, for each test run. Since the control package contained 48 tubes, the sample size in each sequential test was a multiple of 96. λ_n was computed from Equation (3) for the cumulative sample, and the boundary of the "region for accepting null hypothesis" was determined from Equation (13). Figure 2 presents the decision regions as shown in Figure 1, with the observed sequential χ^2 values at the 32nd test run, where the decision to accept the null hypothesis was made. As the figure shows, the eighth χ^2 value falls in the acceptance region. However, in accepting the null hypothesis, it is realized that the χ^2 value would fall in the acceptance region five percent of the time if the alternative hypothesis were true — thus causing a wrong decision.

1.4 The Interpretation of Zero Chi-Square Values

In Figure 2, it may be seen that zero χ^2 values occurred in the first, fourth, fifth, sixth, and seventh tests. This was because the number of removals in one of the groups was only one less than the number in the other group; and, due to the continuity correction factor, the difference between the observed and expected values became zero.

For example, the first zero χ^2 came from the contingency tables shown as Table 2. Since zero χ^2 always falls in the "region for accepting null hypothesis," the inclination is to stop the experiment when this value occurs. However, the occurrence of a zero χ^2 is attributable to the use of the χ^2, a continuous distribution, to approximate the hypergeometric distribution, which is discrete.
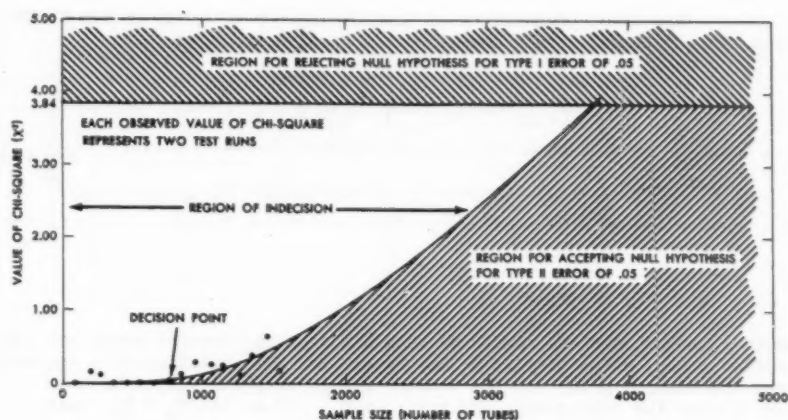
FIGURE 2. PROCEDURE A: SEQUENTIAL CHI-SQUARE VALUES FOR 32 TEST RUNS

TABLE 2. CONTINGENCY TABLES: OBSERVED AND EXPECTED FREQUENCIES

$$\chi_1^2 = \sum \frac{(|0-E| - 1/2)^2}{E} = 0$$

Tube Group	Table of Observed Frequencies			Table of Expected Frequencies		
	Removed	Retained	Total	Removed	Retained	Total
Control	1	47	48	1.5	46.5	48
Treated	2	46	48	1.5	46.5	48
Total	3	93	96	3.0	93.0	96

TABLE 3. CONTINGENCY TABLE FOR EQUATION 14			
Tube Group	Tubes Removed	Tubes Retained	Total
Control	r_1	$n - r_1$	n
Treated	r_2	$n - r_2$	n
Total	r	$2n - r$	2n

437

The exact probability of the contingency of a two-by-two table (see Table 3) based on the one-sided alternative hypothesis is:

$$P = \left[\frac{\binom{n}{r_1} P_1^{r_1} \left(1-P_1\right)^{n-r_1} \binom{n}{r_2} P_2^{r_2} \left(1-P_2\right)^{n-r_2}}{\sum_{r_1=0}^{r} \binom{n}{r_1} \binom{n}{r-r_1} P_1^{r_1} \left(1-P_1\right)^{n-r_1} P_2^{r-r_1} \left(1-P_2\right)^{n-(r-r_1)}} \right] \quad (14)$$

where r_1 and r_2 are the number of removals from the control group and the treated group, respectively; P_1 and P_2 are the probabilities of a removal from the control group and the treated group, respectively; n is the sample size for each group; and $r = r_1 + r_2$. If we let $P_1 = P_2$, then

$$P = \frac{\binom{n}{r_1} \binom{n}{r_2}}{\sum_{r_1=0}^{r} \binom{n}{r_1} \binom{n}{r-r_1}} = \frac{\binom{n}{r_1} \binom{n}{r_2}}{\binom{2n}{r}}$$

which is the general term of the hypergeometric distribution.

When P_1 and P_2 are small and n is large in Equation (14), the Poisson approximation can be used, and P becomes binominally distributed, as shown in Equation (15).

$$P = \binom{r}{r_1} \left(\frac{P_1}{P_1+P_2}\right)^{r_1} \left(\frac{P_2}{P_1-P_2}\right)^{n-r_1} = \binom{r}{r_1} \left(\frac{1}{3}\right)^{r_1} \left(\frac{2}{3}\right)^{r-r_1} , \quad (15)$$

for $P_2 = 2P_1$.

The probability of obtaining a value of zero χ^2 — as illustrated in Figure 2 — was determined from Equation (15); the results are given in Table 4. None of these zero values of χ^2 warrants stopping the test. For example, $r = 3$ in Test No. 1, and

$$\text{Probability of zero } \chi^2 = \binom{3}{1}\left(\frac{1}{3}\right)^1\left(\frac{2}{3}\right)^2 + \binom{3}{2}\left(\frac{1}{3}\right)^2 \left(\frac{2}{3}\right)^1 = .66. \quad (16)$$

438

Test No.	n	Cumulative Removals, by Tube Group		Probability (Two-Tail)
		Control Group	Treated Group	
1	48	1	2	0.66
4	192	7	6	0.20
5	240	8	7	0.16
6	288	8	9	0.14
7	336	10	11	0.10

2.0 ANALYSIS OF PROCEDURE B (THERMAL-SHOCK TEST)

The arc sine transformation was applied to the failure percentages of the "cycle—tube type—manufacturer—environment" experimental unit. An analysis of variance was made, and it was found that cycles did not interact with any of the other factors. Cycles were then compared over the factors. The data are given in Table 5, with a table of expected values. $X^2 = 2.51$, which is less than the $X^2_{.05}$ for two degrees of freedom. Therefore, there is no reason to reject the hypothesis that the three groups have the same removal rates.

TABLE 5. OBSERVED AND EXPECTED FREQUENCIES IN THERMAL-SHOCK TEST

Cycles	Table of Observed Values			Table of Expected Values		
	Failed	Accepted	Total	Failed	Accepted	Total
0	398	521	919	379	540	919
1	345	518	863	355	508	863
3	323	483	806	332	474	806
Total	1,066	1,522	2,588	1,066	1,522	2,588

3.0 ANALYSIS OF PROCEDURE C (VISUAL INSPECTION, X-RAY EXAMINATION, POLARISCOPIC EXAMINATION, VIBRATION TESTING, ELECTRICAL TESTING)

3.1 Bombing/Navigation System

The times-to-tube-removal were assumed to be exponentially distributed. The purpose now is to show that, with this assumption, the statistical techniques used for testing hypotheses and determining Type II errors for the discrete case are applicable here.

Replacements for removed tubes were not included in the computations on this experiment. However, the sample size (approximately 5,000 tubes for each group) was large enough, and the number of removals (about 100 for each group) was small enough, to justify the assumption that the number of removals during the observed total life (μ) followed a Poisson distribution with mean μ/θ. Then, the joint occurrence of r_1 and r_2, given $r_1 + r_2$, is

$$P(r_1,r_2/r_1+r_2) = \binom{r_1+r_2}{r_1} \left(\frac{\mu_1/\theta_1}{\mu_1/\theta_1+\mu_2/\theta_2} \right)^{r_1} \left(\frac{\mu_2\theta_1}{\mu_1/\theta_1+\mu_2/\theta_2} \right)^{r_2}$$

The two samples were compared by the method described in Rao.[5/] (See Table 6.)

$$\chi_1^2 = \sum \frac{(0-E_{H_o})^2}{E_{H_o}} = \frac{(1.9)^2}{124.1} + \frac{(1.9)^2}{118.9} = .05945$$
(Not significant)

$$\chi = .2348$$

$$\lambda = \sum \frac{(E_{H_o}-E_{H_a})^2}{E_{H_o}} = \frac{(5.8)^2}{124.1} + \frac{(5.8)^2}{118.9} = .5540$$

$$\sqrt{\lambda} = .7443$$

From Equation (12), Type II error = 0.147.

3.2 Communications System

The replacements for removed tubes were included in the computations on this experiment. Since an exponential distribution of the times-to-removal was assumed, it follows that there was a Poisson distribution of the number of removals.[6/] (See Table 7.)

$$\chi_1^2 = \sum \frac{(0 - E_{H_0})^2}{E_{H_0}} = \frac{(1.86)^2}{140.14} + \frac{(1.86)^2}{139.86} = 0.0494$$
(<u>Not significant</u>)

$$\chi = .2222$$

$$\lambda = \sum \frac{(E_{H_0} - E_{H_a})^2}{E_{H_0}} = \frac{(18.26)^2}{140.14} = \frac{(18.26)^2}{139.86} = 4.763$$

$$\sqrt{\lambda} = 2.182$$

From Equation (12), Type II error = 0.017.

TABLE 6. PROCEDURE C — BOMBING/NAVIGATION SYSTEM:
OBSERVED TOTAL LIFE, OBSERVED REMOVALS, AND
EXPECTED REMOVALS UNDER THE NULL AND ALTERNATIVE HYPOTHESES

Tube Group	Observed Total Life μ_i	Observed Removals r_i	Expected Removals (E_{r_i})	
			Null Hypothesis $H_0: \theta_1 = \theta_2$	Alternative Hypothesis $H_a: \theta_1 = 1.1 \theta_2$
Control	1,676,710	117	118.9	124.7
Treated	1,749,343	126	124.1	118.3

TABLE 7. PROCEDURE C — COMMUNICATIONS SYSTEM:
OBSERVED TOTAL LIFE, OBSERVED REMOVALS, AND
EXPECTED REMOVALS UNDER THE NULL AND ALTERNATIVE HYPOTHESES

Tube Group	Observed Total Life μ_i	Observed Removals r_i	Expected Removals (E_{r_i})	
			Null Hypothesis $H_0: \theta_1 = \theta_2$	Alternative Hypothesis $H_a: \theta_1 = 1.3 \theta_2$
Control	670,701	138	139.86	121.60
Treated	672,075	142	140.14	158.40

REFERENCES

1   Patnaik, P.B., "The Non-Central χ^2 - and F-Distributions and their
    Applications," *Biometrika*, University College, London; Vol. XXXVI,
    Parts I and II, June 1949; p. 215.

2   Fix, Evelyn, "Tables of Noncentral χ^2," *University of California
    Publications in Statistics*, Berkeley, Cal., Vol. 1, No. 2, pp. 15-19.

3   Rao, C.R., *Advanced Statistical Methods in Biometric Research*, John
    Wiley & Sons, Inc., New York, 1952; p. 40.

4   *Tables of Normal Probability Functions*, U.S. Department of Commerce,
    National Bureau of Standards: Applied Mathematics Series, No. 23;
    June 5, 1953.

5   Rao, C.R., *op.cit.*, p. 205.

6   Epstein, Benjamin, *Outline of Research in Statistics of Life and
    Fatigue Testing*, Wayne University, Detroit, Mich.; Technical Report
    No. 4, p. 7. [Final Report, Contract Nonr-451(00)NR-042-017],
    April 1, 1951 - January 31, 1954.

WHAT DEPARTMENT OF THE ARMY EXPECTS OF CONTRACTORS

Colonel E. J. Gibson, Chief, Procurement Division
Office of the Deputy Chief of Staff for Logistics

Mr. Chairman, ladies and gentlemen.

When Mr. Fales, your Program Chairman, requested Army to help in a session on what Department of Defense expected of contractors, we accepted with the intent of bringing you up to date on Army progress since a year ago when General Cassevant spoke at the National Convention in Detroit and to expand upon the presentation given by General Engler to your Society last fall in New York. Our chairman, Mr. Riordan, has just given a broad statement of Department of Defense policies.

Before telling you of Army policy and how it affects contractors, we should explore differences between the various military departments. Air Force as you know has held strongly to a program of almost complete dependence on their Specification MIL-Q-5923, while the Army and the Navy, except for BuAer, have held out for inspections performed by Government personnel for purpose of determining acceptability. These differences stem from organization patterns and major differences in categories of items which make up the bulk of procurements of the departments.

Like our sister services the Army, in the past few years, has rapidly been getting into complex, continuously changing weapons systems. As a result of these changes the three departments are shifting from widely divergent approaches to inspection and quality control to one which, I believe, in the very near future will be a uniform approach.

This does not mean that any department will adopt the single philosophy of any other department. Rather, it means that the best parts of the programs of all departments will be adopted, for situations where they fit best, by the other departments.

Army has done this. Current Army inspection and quality control policies and procedures are based on a belief in the ability of industry to manage the production and quality control efforts needed to deliver acceptable supplies and services. Basic Army policy was expressed by General Cassevant to your Society last May in his presentation at Detroit. In paraphrase he advised as follows. Contractors are responsible for controlling product quality and for offering to the military departments supplies and services considered by them to conform to contract requirements. Contractors should remain autonomous in the management of their affairs. Army is responsible for determining that contractors have successfully fulfilled their mission. General Engler, last fall, re-emphasized this belief of Army in the competence and integrity of industry.

The revamped Army approach to inspection and quality control is set forth in Army Regulation 715-20, Army Regulation 715-60, the Army Procurement Procedures, and in a Logistics Directive 280-715. The essence of these regulations can be described simply as a requirement that each contractor perform, or arrange for performing, the inspections,

443

including testing, necessary to determine compliance with contract requirements. Army's methods of assuring acceptable supplies and services will be varied as to intensity, depending upon the success of industry in quality control and inspection.

Getting more to the point on what is expected of industry, Army Logistics Directive 280-715 and a revision of Army Procurement Procedures spell out the requirements and how they will be implemented. Suppliers will be required to perform examinations and tests set forth in specifications so as to substantiate conformance of supplies to the specification requirements. In a few instances where suppliers cannot reasonably be expected to perform tests, the Government will so indicate in the specifications or contract, and the Government will perform the tests. These requirements will be spelled out with appropriate contract language. In instances where specifications do not have complete and definitive quality assurance provisions, those specifications will be supplemented in the contract to the extent necessary to clearly define what the contractor must do regarding inspection and testing. Contractors will be required to have available adequate test facilities, or to make arrangements for the utilization of suitable test facilities.

So far I have told you only of what may seem to be additional burdens on contractors. There is a corollary however. Army is re-emphasizing its policy that contractors in all instances should know what to expect in terms of requirements stemming from Government inspection operations. And they should know in contractual form so they can firmly estimate costs generated due to demands of Government inspectors.

These two seeming different missions - one, what would be required of the contractor as a minimum in his inspection to determine compliance with contract requirements, and, two, what may the Government do in assuring contract requirements have been complied with - are both covered very neatly by the same statements. These are the statements that appear in Section 4 of the Specification, or that should appear there. In cases where specifications require amplification, Army policy dictates that the contract shall amplify the specifications. This is in accordance with the policy I read for you from a recent change to the Army procurement procedure.

From your experience in using and producing to military specifications you are aware that some of them do not have complete statements of required quality assurance procedures. Further, some of them will require changes and amendments in the technical requirements. We are well underway in programs which will bring our specifications into line with the new policies I have been discussing. This is being accomplished under two programs. During the last calendar year we trained better than 300 Army people in the principles and concepts underlying specification requirements, both technical and for quality assurance. Two main points are stressed. First, Army specifications will demand no higher quality in technical requirements statements than that necessary to accomplish the military mission. This policy is taught with a background of the major points which set the stage for the level of technical quality needed. These include: functioning, life, interchangeability, weight where it is important, the necessity for conservation of critical materials, and other of the special circumstances which make

military equipment so different in some respects from your normal commercial items.

The second point has to do with contractors knowing what to expect in terms of requirements stemming from Government inspection operations. Army specification writers and engineers are taught in terms of specifications establishing a limit on the maximum severity that the Government may impose through its inspection and quality assurance procedures.

This last policy, even though an Army policy for some time, takes on added significance now that we are requiring contractors to perform inspection in accordance with specification quality assurance procedures. To cope with the need for standardization in the way these procedures are established, Army is preparing an Inspection Engineering Handbook. This handbook, now in finished draft form, is being coordinated throughout the Technical Services of the Army. The handbook defines the basic concepts of inspection engineering and provides the principles and practices associated with quality assurance. It also covers all materials of immediate interest to inspection engineering organizations. Its contents include organization and mission of inspection activities, the meaning of and activities embracing quality assurance, the development of quality assurance provisions, the preparation and maintenance of specifications, equipment lists, standards, and handbooks, and feedback of information.

Another document currently under preparation by Army is a textbook setting forth Army teaching on the principles and concepts underlying specifications. Both of these documents will be available for sale through the Government Printing Office some time during the last half of this calendar year. Some of you who work closely with the military system may be interested in gaining a better understanding of Army basic thinking on specifications.

You may ask what do these changes in policy all add up to. The answer is simple and is almost a restatement of the creed of your American Society for Quality Control. Since little can efficiently be done about the quality of product after it has been produced, then clearly, sorting type operations by military inspectors can only be wasteful. We must push control of quality as far back in the history of items as we can. A first step, of course, is to put responsibility back with the contractor. We do not intend nor are we going to stop with this step.

We have initiated a campaign to do all we can within our persuasive powers to foster within industry greater use of quality control techniques throughout the process of producing or building equipment and facilities for the Army. This is the only way an effective industry-Army team can operate in this modern age. Of course, it always has been the only way. But pressures in the past on our national economy have not been as severe.

In organizing the Army-industry team to get maximum pride of craftsmanship in the production of military equipment, we are seriously considering what may be an about face in policy. We feel that it may be desirable to allow, and possibly sometime in the future to require, all contractors to brand with their commercial trademark the military equipment they produce. This would bring home to all people in our production

445

system the fact that their excellence, as craftsmen, will be on trial all through the life of the military equipment. This, of course, is somewhat controversial at the present time. For instance, how far into the componentry of an equipment will trademarking, and the use of only military trademarked replacement components, be carried?

Another question you may ask, since in the initial stages full-fledged quality control systems seem to be expensive to the producer, how about the inequities that will result when one producer installs and maintains a good system and his competitors do not. The military, and in fact the whole of the United States Government, has recently adopted the general policy-type answer to this question. This is in the form of a change to Standard Form 32, the General Provisions for Supply Contract. Paragraph 5c of the standard provisions has been revised in one place. I will quote: "The Government reserves the right to charge to the Contractor any additional cost of Government inspection and test when supplies are not ready at the time such inspection and test is requested by the Contractor or when re-inspection or retest is necessitated by prior rejection." The revision is in the last ten words of the sentence. Of course, when a contractor does not maintain good quality control and inspection systems, his rejection rate is going to be higher. If, as stated in my quote, he is charged for reinspections or retests made necessary because of rejections, his competitive position will be in jeopardy.

In telling you about the change to Standard Form 32 I used the words "general policy type." These were necessary since much work and study will be required before we are able fully to implement the intent of this contract language.

Everything I have had to say so far adds up to one statement of mission. We are convinced the only way to effectively produce the equipment necessary for our military programs is to so motivate the producers that the supplies they produce will all be satisfactory. We cannot, with Government personnel or industry personnel, rely upon a sorting program after faulty work has been done. Again I'll say, this of course is no more than the creed of your Society.

The Army in striving for this motivation, which requires motivation on the part of Government personnel as well as contractor personnel, has first considered two broad categories of purchase-production systems. The first deals with items that may be clearly described and with which we have had experience in quality control and inspections. The second deals with items complex in nature and so new that concrete descriptions and the details of quality control and inspection are beyond our reach.

For the first category of items we will have clear-cut specifications, in many cases describing design details, and in all cases, establishing the inspection procedures to be used, and will require industry to perform the necessary inspections. In these cases, Army inspection will be reduced to a minimum as made reasonable by the quality of industry's operations.

For the second category, we are striving for the same reduction in Army inspection but since we cannot establish the details of design,

446

nor of the inspection procedures, we will describe contractually the quality controls in general form that contractors must exercise. This will be in the form of a specification MIL-G-14461, titled "General Quality Control Requirements." One of the reference documents in this Army specification will be a Military Standard to be titled "Evaluation of Contractor Quality Control Systems." An interesting analogy may be drawn at this time. As I told you earlier, Army is convinced that contractors in all instances should know what to expect in terms of pressures arising from Government inspection operations. This is true not only in inspection of supplies but also in determinations of contractors' compliance with quality control requirements. Thus, our military standard.

We are publishing this evaluation procedure as a standard rather than an integral part of a specification in Section 4, because we will use the standard administratively to determine eligibility when contractors are producing simplier items for reduced Government inspection.

In summation, the Army expects contractors on all types of equipment to perform the inspections and tests necessary to assure satisfactory equipment. Army will assist industry and expect the assistance of industry in developing more effective quality control programs for reduction of costs and conservation of critical man-hours and materials. I am sure that with the help of the members of this Society the effectiveness of quality control of Army equipment will be vastly improved during the next few years.

447

COST REDUCTION THROUGH CONTROL GAGING

Louis O. Heinold, Jr.
Federal Products Corporation

We are all affected by changing fashions. Whether it is the bill for a new spring outfit, or whether it is the latest piece of precision equipment purchased for our shop, in one way or another we are affected by fashion. A few years ago the word of the day throughout industry was, "automation." Most advertising urged us to use products because of their ability to perform some small task in our own automation program. This urge to join the fashion parade came from manufacturers of hydraulics and pneumatic devices, chucks, pliers, screwdrivers, and virtually any component which could conceivable play a small part in metal fabrication.

Although "automation" is no longer the word of the day, the basic concept which the word implies is still a dominant force in our high production economy. It is no longer fashionable to talk about automation because this is the day of missiles. Today's advertising must tie in with one or more of the many missiles now being manufactured in this country. Whether you are concerned with the manufacture of bearings, electronic components, packaging or office equipment, you must show legitimate connection between your product and a missile when you sell it. A quick glance at any technical periodical will tell the story.

One subject that is always fashionable is Cost Reduction. Today's economic pains make the subject more timely than hitching your product to a missile. And Cost Reduction is the subject of this discussion: How the average small lot producer in the metal working industry can take advantage of some of the latest gage techniques and show some honest to goodness cost reduction at the same time.

We have been seeing many changes in the application and use of gage equipment. Some of these are fundamental changes which started before the war. Naturally during this long period of time there have been many minor trends and side excursions in gage technique and Design. However by recognizing the basic trends it is now possible for the small lot producer to take advantage of some of the cost savings made available by these changes.

Before we go further, let's define the small lot producer. We've been told that about 75% of those engaged in metal working in this country have production quantities of 50 pieces or less. This means that any new Gage techniques which can show signifcant cost savings must have appeal to these small lot producers.

There are two fundamental changes that have been occuring in the application of Gages since the War, and they both arise from the same general development. Specifically we are witnessing an important change in the time and place of inspection. This is the fundamental development of the past years which is changing our gage concepts and should provide more practical and economical inspection techniques to both the small and large lot producer.

Traditionally inspection has always been a "post mortem" tool used to select the good and reject the defects. Today it is becoming a part of the manufacturing process. The time and place of gaging is changing. This basic change is two fold. We are now putting gages on or in

449

the producing machine itself. We are also placing them adjacent to or at the exit of a machine tool.

The result of this trend is to do the gaging in the manufacturing process with the aim that the process can be halted or modified before defective parts are produced. This implies a control function and it also implies that there must be some process that is controllable. One immediately thinks in terms of fully automated production line with a closed loop feedback system in each step of the process. However, because this type of automatic equipment and control system is not useful to the small lot producer, this producer must go back one step further and take a look at the machines in his shop which are controllable either directly from gage information or indirectly from gage information given to a machine attendent.

Now we all know that there are inherent variables in the production output of any particular type of machine tool. In-process gages are expected to provide some kind of control over these variables. To do this it is necessary to understand the kinds of variables affecting the process. Generally speaking there are two types; piece to piece variation and drift variations.

All machine tools, no matter how precisely made, have a certain amount of "shake" or "looseness". This is inherent because the manufacturer has tolerances on the machine tool components. This "shake" or "looseness" will produce some variation in the output of the machine. Likewise, variation in the machine output will occur from deflection of the machine itself. Obviously, no matter how much mass or strength is built into the machine tool framework, the forces required by metal cutting cause deflection of the tool. This, in turn, contributes to variation in the output.

The product variation which occurs because of inherent "looseness" and deflection is called "piece to piece" variation or "scatter."

When the output of a machine tool is statistically analyzed and a histogram of this output is drawn up, the "scatter" of the machine tool is represented by its β sigma limits. This is called a machine capability curve. It is the common desire of the machine tool builder and its user to keep the limits of this curve as narrow as possible. Scatter variations can be controlled by gaging the product while it is being produced and using the gage information to periodically tell the producing machine the exact size of the part being produced.

During a production run on any machine tool, there are numerous trend changes interacting to produce variations in the product. For example, ambient temperature tends to increase during the course of a summer day. Energy released as heat during the cutting action contributes toward product variation. These temperature changes have long-term effects. Some are addative, some nullify each other. At the same time other trend changes in the form of tool wear contribute toward product variation

The net result of these changes are drift variations in the product. They are predictable and controllable by measurin g the product immediately after it is produced and comparing it with preceding pieces.

By moving the inspection function into the manufacturing process we now ask the gage to cycle the machine tool, analyze its output, accept the

450

good and reject defective parts. With gaging in this new role there has to be an economic pay-off. Obviously any gage equipment which is expected to perform such a variety of duties is going to be more costly than our "final inspection" plug gage or dial snap. For any size producer this higher cost has to be justified. It might seem more difficult for the small lot producer so let's look at some of the gains available to him.

The easiest type of machine on which the small quantity manufacturer can control "scatter" is the plain grinder. This is done by providing gage equipment which can measure the product as metal is being removed. The gage information is then used to control the machine. Essentially, such a gage replaces a conventional time base cycle, provides closer control, and a narrower scatter band. Illustration (1) shows a typical installation. The conventional Arnold grinding gage mechanically engages the part. Motion of the plunger of this gage is transmitted to an Air-Probe, which in turn operates an Air Electric Switch. The signal from the Air Electric Switch is amplified by a simple Electronic Signal Unit whose relays are tied directly to the controls of the machine. In this way gage information is given directly to the machine tool.

When control gages are supplied with new grinders, the machines are modified to have two rates of wheel infeed. The normal action of the grinder is rapid wheel advance until the wheel reaches the grinding position. Then wheel infeed proceeds at a fast rate until the control gage signals that the product is about .003" oversize. When this information is received by the grinder, it immediately switches to slow infeed. This rate of wheel advance is designed to replace the usual "tarry" which serves to produce good finish and true roundness. When finish size is "seen" by the control gage, the wheel head is immediately retracted.

Machines which the small-lot producer has in his shop do not have these refinements, but he can still have all the advantages of the technique. Normally, a grinder with a single rate of infeed has an adjustable stop to limit the forward motion of the wheel. Without a control gage, this is set to stop the wheel when the part reaches finish size. The wheel is then permitted to "tarry" or "dwell" against this stop for a pre-set time. This "dwell" produces the characteristic "spark out" and assures the machine operator that the product will be round and have a good finish.

With a control gage on such a machine, two rates of infeed are obtained by setting the stop at a point where the product will be .0002 - .0005" oversize. The timer is set for an abnormally long setting. The action of grinding requires substantial forces. These forces tend to deflect the product and the machine tool. In this setup, the wheel comes in at the usual fast infeed rate until it hits the stop. The product and the machine are both deflected at this point. This deflection diminishes as soon as the forward motion of the grinding wheel is limited by the machine stop. This diminishing deflection literally provides a second rate of infeed. The piece being ground tends to straighten and "spark out". It is rounded up and fine finish is achieved. Its diameter continues to be reduced so when finish size is reached, the gage sends the wheel head back. The timer now acts only as a safety device in the event of gage failure.

This is indeed a good example of how post-mortem measurement can be removed from the inspection area and made a part of the production tool.

451
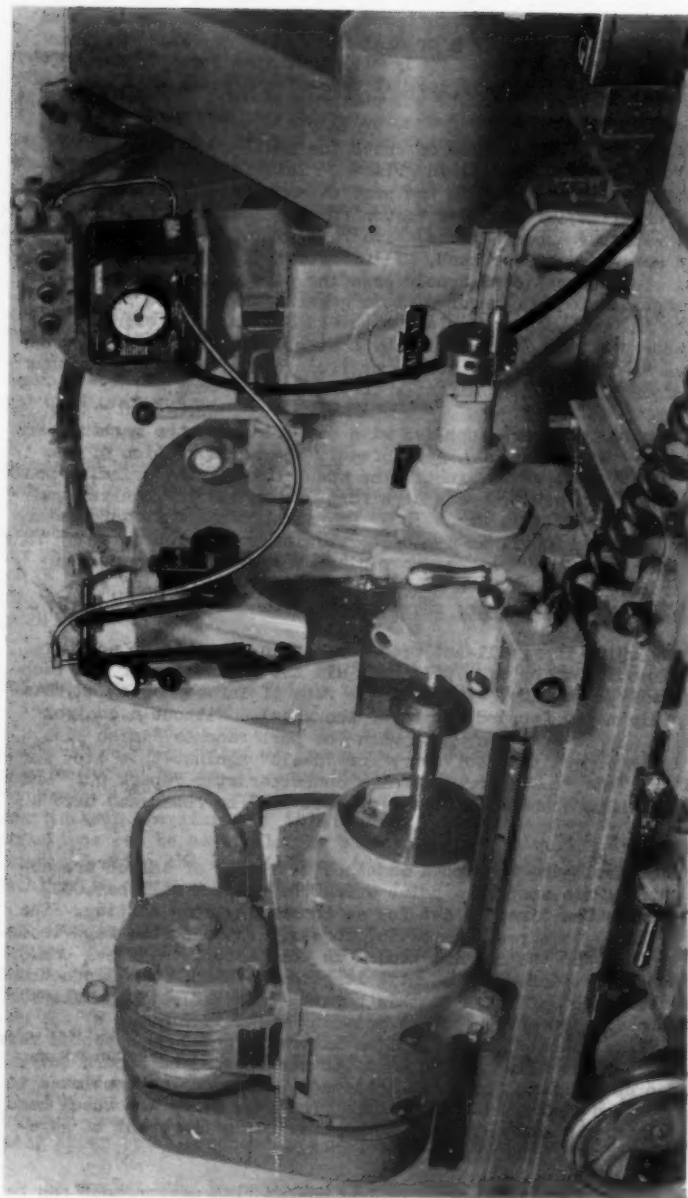
FIG 1 – "In-process" control gage on grinder with gaging caliper in retracted position. Control unit is at top right. Air filter-regulator at lower right. Heavy duty switch unit not shown.

Photo, courtesy of Cincinnati Grinders Inc.

Coupled with a periodic correction for drift variations, parts machined in this fashion should not have to go through another inspection. Time and again, closer control of the product has been proven. On a machine whose machine capability is 60 - 70% of the product tolerance, uniform production in the order of less .0001" variation in size is frequently obtained.

The Arnold Gauge is equipped with adjustable calipers so change over from one size to another is only a matter of a few minutes. This is of tremendous advantage to a small-lot producer. The fact that the operator does not have to be in attendance during the grinding cycle makes it possible for him to supervise two or more machines simultaneously. Here is where real savings can be measured and skilled help freed for other duties.

	MEN	MACHINES	COST PER PIECE	PRODUCTION RATE pieces per 8 hr. shift
BEFORE			$.035 85	1344
AFTER			$.015 26	1386

CONTROL GAGING INSTALLED

SAVINGS (per piece) . . . . $.020 59
RATE INCREASE (per shift) 42
TOTAL SAVINGS (per shift) — $28.59
EQUIPMENT COST AMORTIZED —— 60 DAYS

The chart in illustration #2 shows the tremendous savings which can be obtained by the use of grinding gages on adjacent machines. Here a small compressor crankshaft is ground on two center type plain grinders. Before the use of control gaging, two operators produced 1344 pieces in eight hours at a unit cost of $.03585. Then both machines were equipped with the type of in-process grinding gage we have just described. The machines were physically repositioned so they could be attended by one operator. The new rate of production showed a slight increase to 1386 pieces, but the unit cost was slashed to less than half the old cost. The new unit cost is $.01526 and the equipment cost was amortized in 60 days!

There are additional savings not even shown in this chart. For example the substantial gains in scrap reduction and product improvement haven't been included.

In another case an unexpected additional cost reduction was obtained. This machine, equipped with a control gage had poor hydraulic response. Because of this the first few pieces ground after the machine had been down overnight, or after it had been stopped for operator personal time, tended to be slightly oversize. A piece removed from the grinder and found to be .0001" or .0002" oversize in this case would almost always be scrap. Because there was no accurate way to advance the grinding wheel by such small increments, the removal of such small amounts of stock made salvage impractical. However, by use of the Control Gage and observation of the meter hand only, the operator is able to salvage such parts and show some additional dollar savings for his employer.

These are only two illustrations of how control of piece to piece variation can provide dollar savings to the user. In the average shop, a little imagination will disclose other opportunities for similar control. For example, one manufacturer obtained excellent control results by providing automatic cycling of his pivot polishing equipment.

We have discussed piece to piece variation and how it can be controlled by a gage on the machine which observes the product as it is being produced. Similarly trend variations can be controlled by observing the product immediately after it is produced. Changes in the size of a product can be automatically analyzed statistically in the true quality control fashion. For example, gage equipment measuring the product directly after the producing machine can be set up with four limits of control. Two of these limits are the tolerance limits which sort the piece into "good" and "bad". The other two limits are the inner control limits sometimes called "approaching undersize" and "approaching oversize". These limits are energized only if a sequence of pieces crosses a limit. In some cases, gages of this kind are set up so the inner control limits are energized if two pieces in sequence cross this limit.

The gage shown in illustration #3 provides this type of control of the trend variations of an internal grinder. Gage information is sent directly back to the machine tool which has been designed to accept this information and take corrective action upon its receipt. A machine tool like this has to be of recent origin and specifically designed to function in this manner. Most small lot producers don't have such equipment in their shops and new costly purchases might not easily be justified. However it is possible for this manufacturer to make use of this technique without the costly controls.

One of the most practical applications of gaging after the machine is shown in Fig. 4. This gage is designed to measure the product of a centerless grinder. For many years users of production centerless machines have recognized the advantage of measuring their product at the exit of the machine. Gagemakers have made numerous attempts to solve the problems of such measurement. A severe environment makes a complex mechanical mechanism undersirable. Contact gages indicating product trend caused more problems than they solved. For example, on a machine removing a small amount of stock, a contact gage would put a drag on the workpiece, backing it up between the wheels. This sometimes caused a wheel accident and always affected product finish. The presence of grinding grit made measurements doubtful and generally contact measurement has not been successful.

An air gage is the most logical measuring system to use under these conditions. However, present day air gages have a relatively slow re-
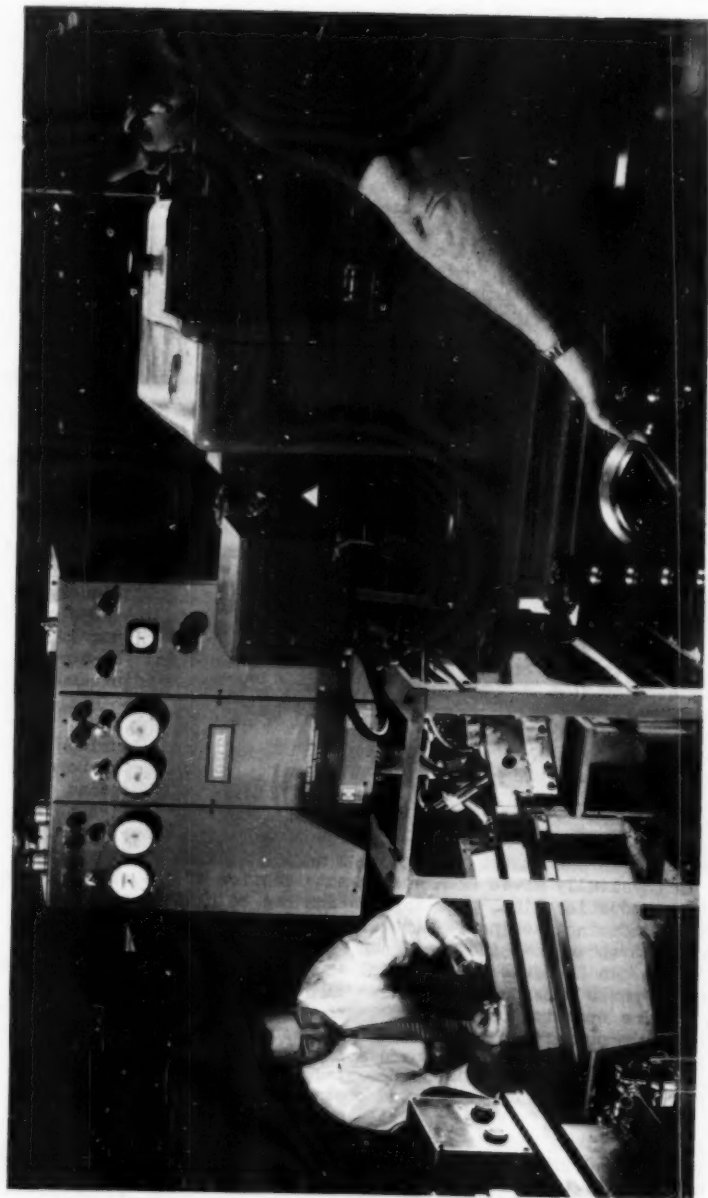
FIG 3 — "Post-process" control gage. Piece rolls down chute (right to left). Air plug connected to control panel above measures I.D. Gage stops machine when size trend approaches tolerance limits.

Photo, courtesy of Link Belt Co.

sponse time. On a centerless grinder they can measure only pieces which
have enough length to satisfy their long response time. Also, undercuts,
holes, or spaces between parts compound the problem of using air gages.

The gage in our illustration (4) meets the important requirement of
non-contact gaging. It is a jetted air ring so there is no drag on the
parts moving thru the machine. Most important, it makes use of an en-
tirely new air system with a response time so fast that high speed grind-
ing of very short pieces is easily accomplished. In one case it is being
used to measure a 3/8" long piece at the rate of 480 pieces per minute.
Similar tooling in the form of adjustable air snaps are now available.

But what of economic justification? Obviously this application is
useful only for production quantities which might be considered optimum
for a centerless grinder. With "after-the-machine gaging" supervising
the trend changes of centerless grinder, it is possible for one machine
attendant to be responsible for a larger number of machines than hereto-
fore. If the current practice is to have one attendant per machine, the
use of control gaging can change this to a two to one ratio. If the pres-
ent ratio is two machines per attendant, it can now be three or four ma-
chines per man. For those manufacturers who use only one such machine in
production, substantial savings can be made by pairing it with a center
type machine equipped with control gaging. One machine attendant operates
both units. We could easily show additional substantial savings just by
picturing the "old method" compared to this new technique. In a shop
where all products go through final inspection, such a department has
personnel, floor space, and equipment. At the same time, setup gages and
possibly the services of inspection personnel are needed on the floor
periodically If a quality control approach is used, it still involves
personnel, equipment, and the need for floor space. These are some of
the things which are drastically reduced when the inspection function is
moved to the producing machine. Add to this the savings made by scrap
reduction, and you most certainly more than meet the basic requirement
of such changes: COST REDUCTION.

The internal grinder equipped with a control gage was described
earlier. This was an automatic installation and probably too elaborate
for the average small-lot producer. To take advantage of the gains ob-
tained with this technique and the same kind of technique applied to the
centerless grinder, it is not necessary to have an automatic process.
With a little imagination, it is very easy to see that if the inspection
function is physically moved from some remote section of the shop direct-
ly into the production line, gage information can be used to adjust and
control the producing machines. Virtually, any well-designed gage, if
placed adjacent to the machine tool, can perform these functions. Inner
control limits can be marked on the dial and machine correction made when
a sequence of parts crosses one of these limits. It is most important
that the machine operator be educated in the new technique. It is point-
less to have him go through the motions of measurement without understand-
ing its significance. He must be made a partner in the new setup. His
cooperation and understanding is fundamental to its success. To a large
extent, final inspection and its extra costs can be safely eliminated.

We have described today's new role for gages. The gages themselves
have to be designed differently to meet these changed conditions. Literal-
ly, the buyer tells the gagemaker, "We expect to use your gage in very
severe environmental conditions. It will not have constant attention.
When it does have attention, this will probably be given by those least
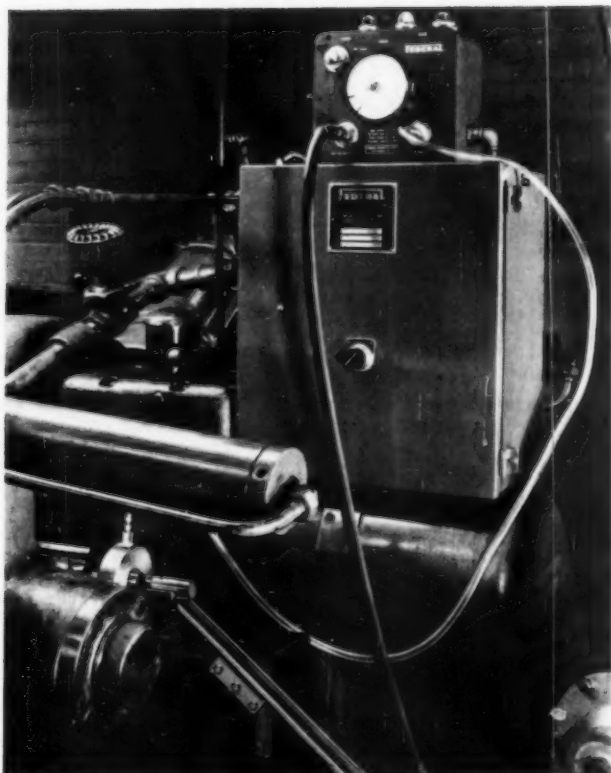
FIG 4 - Centerless grinding gage - continuously measures work at exit of machine, ignoring grooves, spaces between pieces, etc. and warns operator when size trends near tolerance limits.

skilled in its use. We expect it to run for very long periods of time
without drift and without resetting. We want it to perform its tradi-
tional function of separating the "good" from the "bad" and at the same
time help to automatically adjust producing machines." This is indeed a
tall order for the gagemaker. He can meet the requirements, because he
already has recognized the need for new devices, new techniques, and new
systems of measurement. But what about the buyer? Has he recognized
the need for new standards?

The small-lot producer and the large-lot producer too have fallen
behind in the recognition of these new requirements. The self-styled
Equipment Engineer who has been filling his plant with traditional
equipment is faced with a new problem, and in many cases, has done little
to prepare himself for it. Traditionally, Equipment Engineers, Gage En-
gineers, and others have selected gage equipment on the basis of features
which were acceptable for a bench gage, constantly attended and easily
replaced. Great emphasis has been placed on "ease of reading, dead-stop
hands, color, shape". These are NOT the standards for control gage se-
lection. They are being used far too frequently today. They have no
place in the standards which the modern-day Gage Engineer should set for
himself when selecting Control Gaging. Today, the gage is intimately
tied in with the machine tool, gage down time means machine down time.

If the Gage Engineer or whoever is responsible for new gage selec-
tion is to obtain the gains which we have described earlier, he must be
far more careful in his choice of measuring equipment than in the past.
The gages are being asked to do a lot more. It is up to the Gage En-
gineer to be sure they meet the new and higher standard.

If the gage maker were asked to define what this new standard
should be, there would be general agreement that the buyer should select
on the basis of the "best system". This means he has to look past its
cover plate and learn something about the measuring system itself, its
capabilities, its limitations, and what he might expect from it.

If the gage maker were asked to summarize in one word all the rules
which should apply to this new kind of gage selection, again there would
be agreement that the one word is "Reliability". Now this, too, is be-
coming somewhat of a catch word, a "word of the day", so it needs a lit-
tle explanation to see how the overall needs of a Gage Engineer can be
defined as "Reliability"

We have said that such things as bench space, color, speed of hands,
etc. are standards of selection no longer applicable to today's needs.
To arrive at a new standard, the Gage Engineer should select equipment
with a check list of his needs in mind. This check list can be his
definition of reliability. It can help him decide the "best system" for
the job.

Let's take a look at some of the questions that should be on the
Gage Engineer's check list:

1. IS THE MEASURING SYSTEM STABLE? Does the air or elec-
   trical system have the inherent capability of drift-
   free operation. Many of today's air systems drift
   quickly from inital settings. This might have been
   acceptable at a bench setup with an operator in constant
   attention, but such systems make inadequate control

gages. The new gage must be able to hold its setting
without drift for long periods of time. There is
enough drift in the average machining process without
buying more in the form of a gage! As an example, the
Gage Engineer now has to analyze and compare, let us
say, the old-style series air gage and the newer de-
signed balanced air circuit. He has to study electrical
stability and decide if he wants a high current system
with its possible arcing and pitting of gage switch
contacts or a low current system whose initial cost
might be higher, but which would eliminate these unde-
sirable features.

2. <u>IS IT A SIMPLE DESIGN OR SYSTEM</u>? Numerous and complex
adjustments are difficult to make on a gage which has
now become a somewhat inaccessible machine tool com-
ponent. Mechanical adjustments should be few in number,
easily reached, and lock positively. Zero setting of
the measuring system should be no more difficult than
setting your watch. Multiple adjustments for zeroing
involving extra masters (more than one) for setting
calibration and zero simultaneously are time consuming
and deserve no place in today's control gaging.

3. <u>DOES IT HAVE THE REQUIRED SENSITIVITY</u>? The traditional
rule of thumb is that a gage should repeat to 10% of
product tolerance. Product tolerances on centerless
and center type grinding are often .0001 - .0002".
Control equipment adjustable for various sizes has to
be chosen to meet the requirements of the closest toler-
ance. In this case, the entire gage must be sensitive
to .000010". On a test piece under the same conditions
each time, it must faithfully reproduce its reading to
this accuracy. The Gage Engineer should make this re-
quirement a part of his purchase specifications.

4. <u>IS IT STURDY</u>? The supporting gage structure should be
designed to be consistent with its required sensitivity.
It should be strong, but without great mass or bulk.
The measuring system itself should be examined for
sturdiness. In this case, our definition requires
the system be studied for freedom from friction,
hysteresis and other undesirable characteristics of
old-style gages. The entire unit must be examined
for its ability to operate successfully under severe
conditions in the presence of water, grit, and crud.
Check the gage for seals designed to prevent contami-
nation from entering precision bearings and other
important parts.

5. <u>IS IT A COMPLETE PACKAGE</u>? Frequently proposals leave
out units like air regulators, filters and traps.
These are an important part of the proposed gage unit,
which unreliable vendors later tell the customer he
is expected to furnish himself. Other misleading pro-
posals omit the fact that the machine solenoids cannot
be tied directly to the gage switch contacts or gage
relays. Additional electronic or electrical circuitry

459

is required. This means more engineering and additional purchases for the gage buyer. He has been sold only gage components presumably at a lower price under the guise of a complete package. The Gage Engineer should specify that the Control Equipment is complete and ready to tie in to his machine tool without additional re-engineering or purchases.

6. HOW ABOUT SERVICE AND MAINTENANCE? Obviously, a gage which meets the first five tests is not going to require frequent maintenance. Nevertheless, the unexpected may happen so the Equipment Engineer should be sure Service Engineering help for the proposed gage is easily available. The design itself should be checked for easy interchangeability of standard components. A stock of recommended spare parts should be purchased with the gage.

There are other important details which the individual engineer might want to put on his own check list. The important thing is to recognize that this is a new field of gage application. It requires different gage techniques and the ability on the part of the Gage Engineer to scrutinize design, air circuitry, electrical circuitry with complete detachment from his past prejudices. This isn't as difficult as it sounds because most reputable gage manufacturers are very willing to furnish factual advice about their equipment. This, together with a little courage and common sense, will enable the Gage Engineer to make these tests of Reliability a part of his purchase specifications. He will no longer be dependent upon the salesman's glib tongue or misleading advertising. The frosting will be gone, but he will know that he has taken a big step in obtaining the gains we have described. He can assure his management that the end result of his gage selection will be: COST REDUCTION.

QUALITY CONTROL IN THE ASSEMBLY AND MANUFACTURING INDUSTRY
(Abstract)
August B. Mundel
Director of Engineering
Sonotone Corporation

Because a factory makes product which human beings use the product must always be designed and manufactured so as to have the greatest appeal to the end market. The same is true of a management procedure. Quality control is a procedure used by human beings to achieve a result that human beings want. We have a wide range of training among the workers and the supervisors in any industrial establishment. The best results are usually obtained when an operator is given tools to fit his skill and a task within his capabilities but sufficiently difficult to give him some pride in performing the task.

Our quality control jobs are done by people with all types of training and skills. It is a management job to fit the skill and the task. As a result, we often have similar jobs being done in different fashions in different operations.

My discussion will center about some application of quality control procedures that have occurred in our operations and to examine them for the salient points. This is not a discussion of new ideas nor is it the usual application story. It is a clinical discussion of the methods used to achieve a definite end. We have decided to take a group of these applications and discuss them to determine whether they are sound or whether they offer further advantages over systems and procedures more commonly used.

Our quality control organization consists of four sections:

1. Incoming Materials Control.
2. In Process Control.
3. Outgoing Materials Control
4. Experimental and Advisory Service, serving Manufacturing, Research, Enginering and Quality Control.

The four activities cover a similar range of jobs to those described by Dr. Feigenbaum even though the descriptions are not idential. Dr. Feigenbaum lists his four sections as:

a. New Design Control.
b. Incoming Material Control.
c. Product Control.
d. Special Process Studies.

We will use descriptions employed at Sonotone since it forms the basis for the problems to be discussed. The clinical problems are associated with the four divisions. They are:

I(a) The choice of a suitable value for the Acceptable Quality Level, a problem common to incoming materials as well as process and outgoing materials control.

I(b) A discussion of alternate methods of accepting a lot of material coming in in a stratified form.

II. An evaluation of process controls where many similar lines are operated or similar procedures are in use. Can the process all result in similar shrinkage values or similar standard deviations.

III. Methods of maintaining process averages at outgoing material control. In some instances, we cannot afford the consequences of going to tightened inspection and must be most careful to avoid a poor process average. When the process average approaches an out of control condition, what procedure ought to be adopted.

IV. A description of some plant experimentation designed to improve quality of the product or reduce shrinkage (the fraction of product not in accordance with specification).

I(a) The Economical Quality Level (EQL)

If it costs more to do an inspection operation than can be saved, it is economically unsound to do 100% inspection. Our quality control group must determine what quality level is acceptable on an economic basis. This is done on incoming materials and on in-plant transfers by equating the cost of 100% inspection to the savings made by finding and removing the defects. Unfortunately, the economics are not so easily determined on final product. We reach a break even point when the savings equal the cost incurred. Under these conditions,

$$p = \left(\frac{B}{A + E + F}\right) = \text{Economical Quality Level, where } p = \text{fraction defective}$$

at the break even point. The inspection of 100 pieces will yield p x 100 defectives.

The expense is B = the cost of inspection a unit; $100 B is cost of inspecting 100 units. The possible savings are the values A, E and F individually in pairs or collectively as may apply where A = the cost of a unit which we recover when we return a defective unit to the vendor. E = the cost incurred in removing or correcting the defective if it is passed through and found at the next inspection. Sometimes much larger than the cost of the piece, sometimes trivial. F = the cost of processing the part to the next inspection. A waste if we process a defective.

If the inspection cost exceeds the saving made by removing defects 100 x B > p x 100 (A + E + F) we should not do 100% inspection. If the inspection cost is less than the saving made by removing defects 100 x B < p x 100 (A + E + F) we should do 100% inspection. The break even point occurs when 100 B = p·100 (A + E + F) or $p = \left(\frac{B}{A + E + F}\right)$

Figure 1 is our standard form to encourage and simplify this computation.

I(b) The Inspection of Material Received in a Stratified Form

A shipment of small tube micas is received. The lot contains 28,000 pieces. The lot is in a number of small boxes all packed in a larger carton. Each small box has the part number, quantity, production date, and the number of the die used to manufacture the pieces enclosed. In accordance with good sampling procedures, it is the practice of the incoming materials group to sort the material into rational subgroups. These consist of boxes made on the same die and on the same date. The product of two dies is always treated as two distinct lots. If the dates

of manufacture are not continuous, the product of two or more dates is also treated as separate lots. After making the separation by dates and dies, each group is treated as follows: A single sample of size dictated by MIL STD 105 is selected and inspected on an attributes basis for missing holes and obvious defects. The material thickness, a critical dimension, is measured and plotted using approximately 100 pieces from the same sample. From this we compute the grand average and the standard deviation to estimate the possible percentage out of tolerance in thickness. In the past we would check thickness by selecting ten small boxes and checking ten units from each box. If the ten values of $\bar{X}$ (ave) and R ranges showed control and were within specification, we accepted the lot as within specification. The grid hole sizes are next checked on 25 pieces and 6 pieces are checked for hole alignment. These acceptances are made on the basis of the average value.

Why should we check a) a large sample for attributes, b) a smaller sample for thickness, c) a still smaller one for grid hole size, and d) an even smaller one for alignment? A good bit can be said for a and d. The attributes sample indicates that the punch and die set is operating and if it is operating consistently it is not likely that a systematic shift in hole location can occur. Hence d must be invariant if the holes are present. Why should two variable b and c use such vastly different sample size? What is the relative efficiency or advantage of the two methods of acceptance of the mica thickness?

III The Control of Outgoing Product Process Average

The product quality record is reviewed on the basis of the last ten lots submitted. The product has been accepted on the basis of normal sampling. Should the process average exceed the upper limit of the process average for the appropriate AQL, we must go to tightened inspection and incur other penalties. This we wish to avoid.

The acceptance has been on the basis of MIL STD 105 single sampling. Lot size 801 to 1300, sample code letter K, sample size 110, AQL 2.5%. The acceptance number is 6. The rejection number is 7. The last 9 lots have accumulated 39 rejected pieces. The tenth lot is about to be submitted. We examined the upper limits of process average (Table II MIL STD 105A) and find for a sample of 1100 pieces (sample of 110 each from 10 lots) the upper permissible process average for a 2.5% AQL is 3.87% or 42 pieces. We do not want to find 4 or more defectives in the next sample of 110. Since our last two lots have had 5 and 6 rejects in the sample, we suspect our process has drifted to an undesirable level. How shall we treat the lot to make sure that when it is sampled the number of rejects is less than 3? Our procedure must obviously be to 100% inspect to tightened limits and then sample the lot to assure ourselves that when inspected by the quality assurance group the lot will yield 3 or less rejects. Is it sufficient to go to tightened sampling inspection or must we be even stricter in this instance? If we must be stricter, how do we accomplish this within the confines of the MIL STD 105 tables. We must assure ourselves by some sampling procedure that the lot when sampled will have a high probability of yielding less than 3 defects in the 110 units checked. Suppose we decide that we can risk a .05 probability of rejection. This, in turn, implies that we must be 95% certain that the lot contains no more than 3 out of 110 defectives. In terms of the consumer risks of .05 probability of acceptance this agrees closely with the 0.10% AQL for the same sample code letter K,

463

size 110 pieces from the curves (Figure 2) of MIL STD 105A. According to the sampling plan for an AQL 0.1 and the same lot size 801 - 1300, the sample size 110, the acceptance number is 0 and the rejection number is 1. This checks well with the Poisson. According to the Poisson, when p = 3/110 (the average fraction defective) and a sample size n = 110 is used, the product (pn) = 3.0. For this sampling plan, using an acceptance number of 0, the probability of accepting a lot of this quality is .05. Only once in 20 times will we accept a lot poor enough to actually have 3 defects in a sample of 110 on the average.

In effect, we have applied MIL STD 105 and have gone far beyond the tightened inspection procedure to achieve a closer control of a specific lot on a lot tolerance % defective basis for a specific purpose.

IV Experimental Designs

The experimental and advisory service works with research, engineering, production and inspection groups, assisting in the analysis of data. Preferably, this group assists in setting up the experiments in order to assure the results.

An example of poor experimental design is represented by the work of one engineer who was advised that the customer reported 10% of the product defective. He was quite certain that the cause of these defects had been eliminated but just to give himself greater assurance he checked a sample of three units and found them satisfactory on an attributes basis. He and his group were chagrined at the possible error after it was demonstrated that the probability of selecting three satisfactory units from a universe containing 10% defective was $.9^3$ or .73. Advisory service before the experiment avoids this type of error.

An example of a factory experiment designed to evaluate the causes of defects in a production line was reported in Inspection Data and A Trouble Shooting Expedition[1]. A combination of two levels of five variables was explored to determine the contribution to fraction defective of assemblers, assembly jigs, adjusters, adjusting jigs and final inspectors. The experimental design is shown in Figure 3. Forty units were processed through each numbered box. Each part of the experiment and each move of product down the line was planned before the experiment was started. Upon completion of the experiment, the work was all collected and held while the analysis of the experiment was done in a most simplified manner to provide an immediate answer as to where the difficulties were and to help educate our personnel on true experimental procedure.

The important parts of the program are 1) team effort to complete a task, 2) recognition and use of the advantages of statistical knowledge to identify areas of difficulty in contradistinction to the earlier example of the misguided engineer's use of too small a sample.

(1) George Scheel, Sonotone Corporation, ASQC Middle Atlantic Regional Conference, February 28 - March 1, 1958.

464

QUALITY SPECIFICATION Q30,073   SHEET 2   ISSUE 1

SUBJECT _____ PROCEDURE FOR COMPUTING ECONOMICAL AOQL

Part # _____ Item # _____ Name of Part _____

Calculation of Economical AOQL

Date					
A. Raw Material or Assembly Value $/Unit					
B. Direct Cost of Inspection/Piece $/Piece					
or C. Inspection Rate/Hour					
D. Inspection Cost in $/Hour					
E. Cost of Removing or Correcting Defect After Assembly $/Unit					
F. Cost of Processing to Next Inspection					
$p = \dfrac{B \times 100}{A+E+F}$ or $\dfrac{D \times 100}{C \ (A+E+F)}$					
$p = $ AOQL (economical)					
Use AOQL of					
Calculation by					

Figure 1

465

TABLE VI-K.—*Sampling plans for sample size code letter: K*—Continued

OPERATING CHARACTERISTIC
CURVES FOR SINGLE SAMPLING PLANS

(Curves for double and multiple sampling are essentially equivalent)

SAMPLE SIZE CODE LETTER: K

Figure 2

466

		A1				A2			
		B1		B2		B1		B2	
		C1	C2	C1	C2	C1	C2	C1	C2
D1	E1								
	E2		3	4	6		1	2	
D2	E1	5				7			
	E2								8

A1 AND A2 ARE ASSEMBLERS

B1 AND B2 ARE ASSEMBLERS JIGS

C1 AND C2 ARE ADJUSTERS

D1 AND D2 ARE ADJUSTERS JIGS

E1 AND E2 ARE INSPECTORS
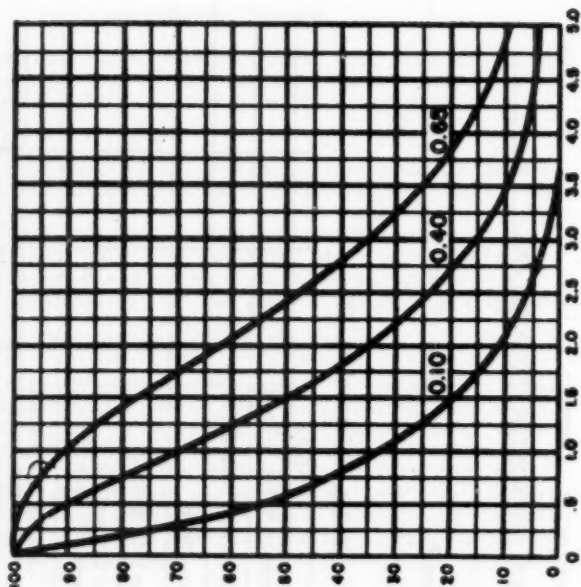
NUMBER IN BLOCK IS TRAY NUMBER AND INDICATES ROUTING THRU OPERATORS, TOOLS AND INSPECTORS. FOR EXAMPLE TRAY #3 WILL GO TO ASSEMBLER #1 USING ASSEMBLY JIG B1 THEN TO ADJUSTER #2 USING ADJUSTING CLAMP D1 AND WILL BE INSPECTED BY INSPECTOR E2.

Figure 3

467

Control Charts Without Calculations
Some Modifications and Some Extensions

Paul C. Clifford
State Teachers College
Montclair, N. J.

Abstract. The usual control chart for variables requires a separate data sheet, a reasonable amount of arithmetic, and two charts. The computation of control limits frequently requires the elimination of some of the original data. And for shop personnel there is still confusion between control limits and specification limits. A method is presented in which individual measurements are plotted, thus eliminating the data sheet. Control limits are established by a process of measuring and counting. Such charts lend themselves to a variety of applications, some of which are considered. In particular this procedure gives a simple comparison of process capability and process achievement.

The Industrial Efficiency of Control Procedures. Control charts using the range have almost completely replaced charts using the standard deviation. One reason is simplicity of computation. A second is that the range is a better indicator of certain sorts of trouble. Considering the industrial usefulness of the range, it is surprising that substitute measures for the average-range chart have received so little consideration. Ferrell in an excellent paper [1] has presented the advantages of the mid-range. Much of the theory in Ferrell's article is pertinent to the charts presented in this paper. The Narrow-Limit gaging procedure of Utt and Mundel [2] and the Pre-Control procedures developed by Shainin and Satterthwaite [3] are examples of plans that are based on the characteristic of an individual item rather than on some statistic computed from these individuals.

There is a wide variety of control problems for which the quality control engineer needs an extensive kit of techniques. Operationally the problems run from first piece inspection and equipment set-up inspection to multiple factor experiments designed to discover well concealed causes of trouble. Statistically we have a battery of tools ranging from distribution analysis and control charts to regression analysis, analysis of variance, surface response estimation and so forth. In evaluating any industrial procedure I would suggest that there are at least four aspects: engineering, economic, statistical, psychological. Just because the statistical aspect is the easiest to evaluate, we tend to overvalue the statistical efficiency and lose sight of the other aspects.

Thus the charts that I will present have advantages, but in general they are not based on arguments regarding statistical efficiency. Rather this is a report on several years industrial experience which has confirmed my intuitive belief that such charts have a wide area of application. In general this area is definable as "Problems or processes that require more than first piece inspection and less than three factor designed experiments." I do not propose that these should replace $\bar{x} - R$ charts. My belief is that there is no universally best chart--the proper one will depend on the process, the personnel, and the type of instrumentation and the importance of the problem.

469

Control Charts Using Individuals in Sub-Groups. To facilitate compari-
son with the usual chart I have taken an example from Grant (4) page 18.
I shall not reproduce the data, for in the proposed procedure, no
record form is used. Rather the readings in each sub-group are plotted
directly, the five points being plotted along a vertical line segment as
shown in figure 1. For this one should use rather coarse graph paper
with large squares, since for most characteristics 10 to 30 units of
measurement cover the range of variation encountered. We now proceed
to change this plot of individuals to a control chart.
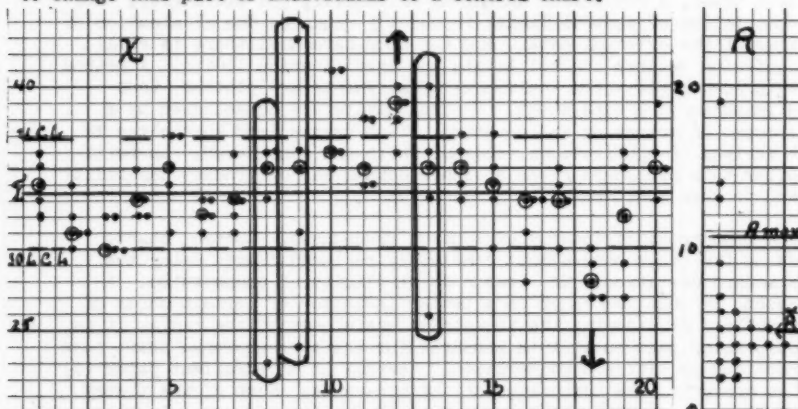
Figure 1

The first step is to indicate the median of each sample, which we
have done here with a box. Next, using a reference card, the indivi-
dual ranges are <u>measured</u>. To do this the bottom edge of the card is
placed at the lowest reading of the subgroup, and a mark is recorded
opposite the largest reading in the same sub-group. Having marked the
20 ranges on the card (shown on the right of figure 1), we count down
to the middle mark. This gives us the <u>median range</u>. While it is
sometimes necessary to interpolate, it is usually better to be conser-
vative, so we take the median range as 5. We next look up the factors
for the control limits on the range. These are given in the appendix
as D_3 and D_4. In this case D_4 equals 2.179. This gives a range con-
trol limit of 10.5. The index card shows that three ranges are outside
this limit. We make a range control rod as shown on the left, with
the median range and the maximum range indicated. We then locate these
three subgroups that are out of control and circle them as shown on
the chart.

Next we draw in a <u>center line</u> on the chart. Here we may use
either the <u>median of the sub-group medians</u> or the <u>median of the indivi-
dual observations</u>. The number of subgroups is the deciding criteria.
With only 20 sub-groups it is better to take the median of the indivi-
dual observations which in this example is 33.5. We then look up the
A_2 factor for medians which for n = 5 is .712. This factor multiplied
by the median range gives a product of approximately 3.5. Control
limits are drawn at this distance on both sides of the center line.
Sub-groups 12 and 18 have medians that are outside these control limits.

470

In a process that is so badly out of control there is not much use in adding limits for individuals. This has been done here just to illustrate the procedure. Dividing the median range by the d_2 factor for n=5 gives an estimate of sigma for individuals of 2.2. Multiplying this result by 3 gives a product of 6.5. Control limits for individuals can be drawn at this distance on either side of the central line. These control limits are 27 and 40. These are our estimates of process capability provided a state of control can be maintained.

At this point it might be pointed out that the usual analysis as given in Grant arrives at essentially the same result. However, to do this requires two or three recomputations as out of control results are discarded. Such procedures tend to make the novice feel that if he keeps at it long enough he will discard all the data.

We may wish to have an overall measure of what the process actually produced. We could of course make a frequency distribution of individuals and find the mean and standard deviation of this distribution. However, such accuracy is seldom needed. we already have 33.5 as a measure of center. What is needed is a measure of variability which includes the five samples that showed out of control conditions. To obtain this simply we count in 7 observations from each end. This gives us the 7 and the 93 percentile points. In this example these are 28 and 38.5 respectively. Taking one third of this difference (the correct multiplying factor is .3388) gives an overall standard deviation of 3.5 as a measure of process achievement. If we compare this overall standard deviation to the estimate of the within group variability (3.5/ 2.2) we find that the variation in the process is approximately 60% more than we would have if the process could be controlled.

Finally, it seems to me that control charts should be summarized before we become embarrassed by ancient history. Let us assume that the above data represents some unit - production, time or merely the end of a graph sheet. It can be summarized as follows:

Average	34.5
Process Capability	± 6.5
Process Achievement	±10.5
X out of control	2
R out of control	3

Evaluation and Criticism of the Median Sub-Group Chart. The chart has several definite advantages that might be listed:

1. The data sheet and the arithmetic involved in finding X̄ and R̄ have been eliminated.
2. Since both the X median and the R median are not greatly affected by out of control data, recomputation of the center lines and control limits is practically never necessary.
3. The center line on the chart is the proper one for runs above and below the median.
4. Since individual results are plotted, it is possible to show both control limits and specification limits.
5. Comparisons between process capability and process achievement are easy to make. The chart can easily be summarized.

Any chart has disadvantages - at least in the wrong application. If the individual observations on the data sheet have auxiliary information tied to them - such as the piece part number, machine position or material - we may find that a data sheet is necessary. If the data sheet is necessary, I see no reason for using this type of chart. It is obvious that the chart emphasizes the average and soft pedals the range, and that it is most useful in those applications where the control of the average is of major importance.

The Mid-Range Sub-group Control Chart. A rather obvious adaptation is the chart where Mid-Ranges are used as the measure of center. In this case we plot only the extreme readings and the average of the extremes. The analysis of ranges is the same as we have outlined above. The control limits for Mid-Ranges require a slightly different factor which I have indicated by $\overline{A_2}$. This chart has many of the advantages listed above. It does not have advantage number 5. I find that the decision between the median chart and the mid-range chart depends to a great extent on the process. There are some generalizations that help in this decision:

1. From the viewpoint of statistical efficiency for sampling from a normal distribution, the midrange is as good or better than the median for sub-groups of n = 3,4,5. For n greater than 6 the median is more efficient.

2. The statistical efficiency of mean, median and mid-range all depend on the type of distribution. The mean is most efficient for the normal distribution. For distributions with limited range the efficiency of the mid-range increases (5). For distributions with extremely long tails such as the Tukey distribution (6) the median is the more efficient.

However, sampling from a fixed population is only part of the story. If we have lack of control we have an indication that there is no one population involved. In such a case the median range is superior to the average range, since the 'contaminated' ranges have less effect on the median. If the percent of contaminated ranges is very high, say 20 or 30 percent, this procedure will be poor, but so will any procedure.

This question of contamination is important in choosing between the two charts. Suppose one sub-group had four typical readings and one unusually high reading. This high one we shall suppose is a 'wild shot.' Do we want our measure of center to show as outside our control limits. I think the answer is that in production we would not, while in research and development we would. I believe that this may be the reason why Ferrell prefers the mid-range chart and I prefer the median chart - we are expressing our differences as to what we want the chart to show in different types of applications. On a sample like the one described I am afraid that the operator will reset the process-- Ferrell is afraid that the engineer will ignore possible troubles.

We have admitted that if the process is in control either the mid-range or the median is likely to be less efficient than the average. However, I question whether this is a real deficiency and might even claim that it is an advantage. I find that the usual trouble with control limits is that economically they are too 'tight' and that as

a result there is difficulty in getting people to take any action when the charts show points out of control. It may be that wider control limits would be a real benefit. As an example consider a process with specification limits of $\pm 4\,6$.

Control limits for the average would be	$\pm 1.34\ 6$
Control limits for the median would be	$\pm 1.61\ 6$
Modified limits for the average would be	$\pm 2.34\ 6$

It appears that the less sensitive limits for the median would not be a disadvantage unless the maximum control is needed.
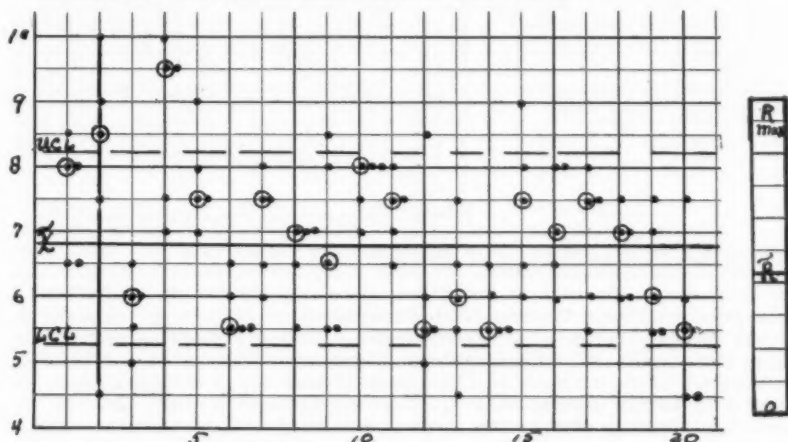
Figure 2

Control Charts for Sub-Groups, Two way classification. In many applications each observation may be classified in two ways. One is usually time, the second is frequently position. As an example figure 2 shows the results of Preece Dip Test on the amount of galvanization on wire. In this case a galvanizing bath carried thirty wires. For sampling purposes the bath was considered as being in five sections, and one wire was selected from each section. The wire numbers of the highest and lowest readings were recorded on the chart, but have not been shown here. Samples 2 and 4 are out of control on the median. The cause turned out to be the method of charging plus a defective electronic control. Sample number 2 is also out of control on the range. Both the high and the low readings were investigated. It was discovered that the same wire was sampled on samples 2, 12, 13, and 16, the readings being 4.5, 5.0, 4.5 and 6. It appears that this wire is giving unusually low readings. The cause appeared to be a defective die.

This example is given to indicate the type of application. I have found this most useful in analysis of multiple spindle machines, filling operations with multiple heads, batch sampling with stratification and so forth.

473
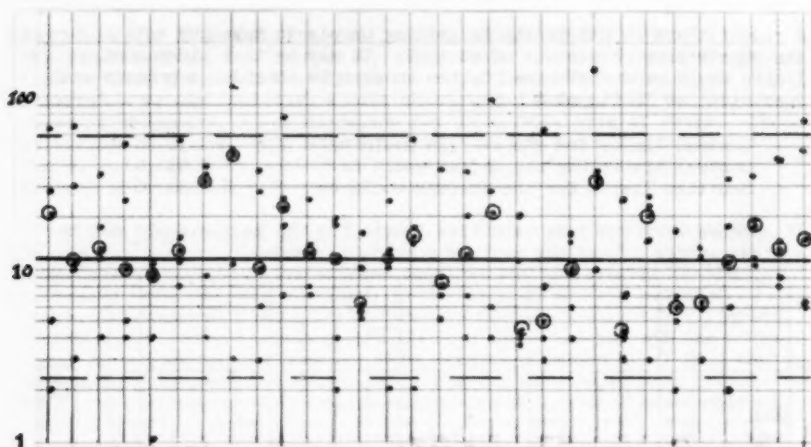
Figure 3

Control Charts for Skewed Distributions. The distribution of the
median is approximately normal for moderately skewed distributions.
For very skewed the data is sometimes transformed by taking logarithms.
This same result can be accomplished graphically by plotting the indi-
vidual readings on semi-logarithmic paper. Figure 3 shows an example
taken from Cowden(9). The individual items are plotted, the median of
each sub-group is circled, and the range of each group is measured
using a reference card as before. The Median range is then found by
counting. In this example the median range is just less than the
length of one cycle on the log scale. This distance is then multiplied
by the factors D_4= 2.179 and A_2= .712. The ranges in this example are
all in control. The distance to the control limits for medians is
measured off on each side of the center line. In the example the cen-
ter line is 11.5 and the control limits are 2.3 and 58. All the medi-
ans are inside these control limits. Three sigma limits for individuals
can be found by taking the median range (distance), dividing by
d_2= 2.257 and multiplying by 3. This distance when laid off from the
center line gives three sigma limits for individuals of 0.5 and 250.

The usual control chart for average and range shows several points
out of control as can be seen in Cowden. Cowden handles this problem
by finding the log of each observation, and making a control chart for
this transformed data. This involves looking up the logs of 150 numbers
plus the arithmetic to make a control chart. The time by the proposed
method is that required to plot the readings, plus three minutes to
find the limits. The results are practically identical.
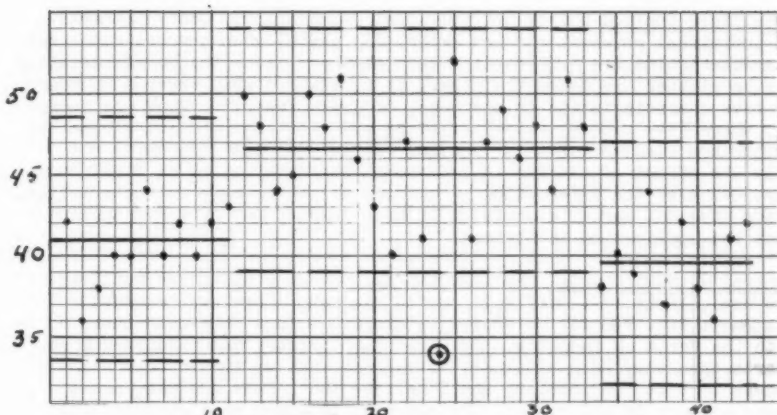
Figure 4

<u>Control Charts for Individual Observations</u> In some applications,
especially in the process industries, there is no rational way for
sub-grouping the data. In other applications it may be that a sam-
ple of more than one is not justified economically. Figure 4 shows
individual test results on a natural product used in the chemical
industry. The ranges are found by measuring from one reading to the
next. The median range is 3.5. Since individuals are so insensitive
to shifts in average we used two-sigma control limits. These are
obtained by dividing the median range by d_2 of .954 and multiplying
by 2. This gives control limits of $\pm$ 7.5. The median of the X's is
42.5 and we can easily see that there are many points out of control.
Examination of the chart shows evidence of either a cycle or a shift
in average. In this case a cycle would have no operational meaning,
but a shift in average could be readily explained.

To analyze the data for shift, a center line and control limits
were drawn on clear plastic. This plastic overlay is then shifted up
and down vertically. The number of runs above and below the median
must be watched carefully. After several trials we find that there
appear to be three different levels or batches, as is indicated on the
chart. We cannot guarantee that the batches terminate where the chart
indicates they do. However if we can adjust the process to compensate
for difference in batches, we have the approximate place to make the
change. One test result on sample number 24 is a "wild shot" and
should be investigated or the material set aside.

The use of any gadget, such as the plastic overlay is difficult
to describe and easy to demonstrate. There is of course the criticism
that one does not have exact probability levels for such an eye test.
The same is true of any similar procedure, such as fitting a line to
cumulative data on probability paper. Moreover I know of no exact
statistical method for segregating batches such as we have attempted
here. I can state that the actual industrial application was success-
ful, but I cannot deny that I may have been fortunate. However the
same procedure is very useful when we know that there are definite
batches.

475

<u>Multiple Control Charts on Individual Observations</u>. Control charts on individual observations are especially useful when several different characteristics are recorded for each sample. Since the characteristics can be plotted one below the other, it is easy to keep records on four or five characteristics on one sheet of graph paper. I know one industry where five process characteristics and two final product characteristics are analyzed in this fashion. It is possible to observe not only the effect of changes in individual characteristics on final quality, but also the interaction in changes of two or more characteristics. Space does not permit discussion of such charts, but it would appear that the application is obvious.

<u>Control without Charts</u> The question as to when a chart has outlived its usefullness is a difficult one to discuss. However the usual statement that removal of charts is dangerous just does not square with the facts of industry. By this time industry has discarded more charts than they have retained. I suggest that we should take a good honest look at the problem, and I believe that there are many cases where a chart is not needed. My question is - <u>How much control?</u> The control chart for medians can be easily adapted to a procedure without charts. To do so requires that the process capability has been determined and is less than the specifications. To be specific let us refer to figure 1. Let us assume that the process is near a state of control, that the target of 33.5 is satisfactory, and that control is to be based on a standard deviation of 2.5. The operator is given the following instructions:

1. Measure each of a sample of five items
2. The median is to be between 33.5 ± 4.0 $(X \pm \widetilde{A} 6)$
3. The range is to be less than   13
4. If the results are out of control in one out of 20 checks the process is satisfactory. If out of control more than one time in twenty, a chart will be temporarily reinstituted.

This procedure can be easily adapted to most measuring instruments by using a slightly frosted plastic overlay. On this the operator marks each reading as he takes it. The control limits for the median are pre-marked on the overlay. If it is necessary a parallel scale giving the limits for the range could be attached. I have never found this particularly useful.

<u>Summary</u> Industrial experience has indicated that there are many cases where a simpler procedure than the usual control chart would be useful. Some of the advantages of control charts using individual observations, with control based on the median, have been indicated. Some applications to different types of control problems have been indicated. These charts have been used in a variety of industries and the economic and psychological advantages appear to outweigh any statistical disadvantages in many applications.

References

1. Ferrell, E.B., "Control Charts Using Midranges and Medians"
   Industrial Quality Control 10, 30-34 (1953)
2. Ott, Ellis R. and Mundel, A.B., "Narrow-Limit Gaging"
   Industrial Quality Control 10, 21-28 (1954)
3. Shainin, Dorian, "Quality Control Without Charts"
   Quality Control Convention Papers, A.S.Q.C., 1954
4. Grant, E.L., "Statistical Quality Control"
   McGraw-Hill, New York (1952)
5. Rider, Paul., "The Midrange of a Sample as an Estimator of the
   Population Midrange". A.S.A. Journal 52, 537-542. (1957)
6. Hastings, C., Mosteller F., Tukey, J.W., and Winsor, C.P.
   "Low Moments for Small Samples", Annals of Mathematical
   Statistics 18, 413-426 (1947)
7. Godwin, H.J., "Some Low Moments of Order Statistics"
   Annals of Mathematical Statistics, 20, 279-285, (1949)
8. Dixon and Massey, "Introduction to Statistical Analysis"
   McGraw-Hill, New York (1957)

Mathematical Appendix

Unfortunately there is no standard set of symbols for medians and midranges. The symbols used here have one advantage - the formulae for median and midrange charts are similar to the standard one for $\tilde{X}$ and R charts. The symbols are defined as follow:

$\tilde{X}'$ = median of parent universe
$\tilde{X}$ = median of sub-group
$\overline{\overline{X}}$ = midrange of sub-group
$\tilde{X}_i$ = median of all individuals in sample
$\tilde{\tilde{X}}$ = median of sub-group medians
$\tilde{\overline{X}}$ = median of sub-group midranges
$\tilde{R}$ = median of sub-group ranges

$$\tilde{d}_2 \, 6' = \tilde{R}' \qquad\qquad\qquad\qquad\qquad \text{Standards Given}$$
$$\tilde{A}_2 \tilde{R}' = 3 6_{\tilde{X}} \qquad \tilde{A}_2 \overline{R}' = 3 6_{\overline{X}} \qquad \tilde{A} = A / \sqrt{E_{\tilde{X}}}$$
$$D_3 \tilde{R} = D_3 \tilde{R} \qquad D_4 \tilde{R} = D_4 \tilde{R} \qquad \overline{A}' = A / \sqrt{E_{\overline{X}}}$$

Estimates : $\tilde{X}' = \tilde{\tilde{X}}$ or $\tilde{\overline{X}}$ or $\tilde{A}$
$$6 = \tilde{R} / \tilde{d}_2$$

3 6 Limits for $\tilde{X}$ $\tilde{\tilde{X}} \pm \tilde{A}_2 \tilde{R}$ or $\tilde{\overline{X}} \pm \tilde{A}_2 \tilde{R}$
$\overline{X}$ $\tilde{\overline{X}} \pm \tilde{A}_2 \tilde{R}$
R $\tilde{D}_3 \tilde{R}$ and $\tilde{D}_4 \tilde{R}$

Note: $E_{\tilde{X}} = 6_{\overline{X}} / 6_{\tilde{X}}^2$. This is tabulated in reference 8. The tables that follow have been computed from references 1, 6, 7 and 8.

Factors for Median and Midrange Charts Using Median Range
Normal Distribution Theory

n	$\widetilde{A}$	$\overline{A}'$	$\widetilde{A}_2$	$\overline{A}_2$	$\widetilde{d}_2$	$\overline{D}_3$	$\widetilde{D}_4$
2	2.121	2.121	2.224	2.224	.954	0	3.865
3	2.014	1.806	1.265	1.137	1.588	0	2.745
4	1.637	1.637	.829	.829	1.978	0	2.375
5	1.615	1.532	.712	.679	2.257	0	2.179
6	1.387	1.458	.562	.590	2.472	0	2.055
7	1.385	1.402	.520	.530	2.645	.078	1.967
8	1.233	1.358	.441	.486	2.791	.139	1.901
9	1.240	1.322	.419	.453	2.916	.187	1.850
10	1.216	1.293	.369	.427	3.024	.227	1.809

TESTS OF SIGNIFICANCE I - NORMAL MODELS

Gayle W. McElrath
Head, Industrial Engineering Division
University of Minnesota

Introduction:

Many of the concepts that are fundamental to tests of significance are already the property of the quality control engineer, and their application involves at most an extension of the principles of statistical quality control. The main effort of this paper is directed toward introducing those statistical concepts and methods which give a quantitative measure to situations for which the quality control engineer already has a strong intuitive feeling. We can only hope to illustrate a few of the tests of significance concerning the normal model in order to give an appreciation of their usefulness.

The average quality control engineer is not making the most of significance tests. However, there are many opportunities in each area of quality control for the applications of such tests. In many instances the engineer cannot answer questions concerning process level or variation because he is not acquainted with—or even aware of—the statistical tools which are available.

NEW WORDS FOR OLD IDEAS

One of the reasons for the lag in time between the theory and methods developed by statisticians and their application by the quality control engineer is simply a matter of language difference. The precise terminology of the mathematical statistician often seems difficult and even "unreasonable" because of its unfamiliarity. However, the quality control engineers who make a sincere effort to understand the language of statistics are early rewarded by finding that the basic and fundamental ideas of significance tests are natural and extremely useful.

The engineer considers "a variable which is always subject to some stable system of chance causes." [1] The outcomes of system of chance causes can be of many types. However, analysis is easier when the outcomes are assigned numbers; we attempt to identify a number with each outcome of the chance cause system or chance experiment. This assigned value is called a RANDOM VARIABLE. We make the following definition.

Definition: A random variable is an assignment of numbers to outcomes of a chance experiment.

A random variable is called "random" because the experiment is an experiment of chance. It is called "variable" because it is a numerical quantity which varies from one trial to another trial of the experiment. The frequency of occurrence of various outcomes can often be represented by a MATHEMATICAL MODEL of the numbers to which probabilities of the outcomes have been defined. Such a model might also be called a FREQUENCY FUNCTION which gives the pattern of the potential values that the random variable can take on.

This paper assumes that the reader is acquainted with the concepts of mean, variance, and standard deviation. In the following discussion the terms mean, average, and process level will be used synonymously.

However, we must emphasize that it is of the greatest importance to distinguish between the mean and variance of a population which are PARAMETERS and the similar quantities calculated from a random sample which are called STATISTICS. A parameter is a population measure which is usually an unknown constant, and a statistic is a sample measure which is a random variable.

When (1) sampling with replacement and mixing, (2) sampling without replacement from an infinite population, and (3) sampling in cases where the population had to be imagined, the observations of the sample are independent, and they share a common pattern of potential values. Furthermore, that pattern of potential values for each observation is exactly that of the underlying random variable or "population." We introduce, then, the following definition:

Definition: A random sample is a sample whose n observations may be represented by the mathematical model of n independent random variables, each a replica of the population from which the sample is taken.

Even though we shall consider principally random samples in the sense of the above definition, we should mention that it is possible to modify the formulas to cover the case of sampling without replacement from a finite population. Most of the theory of statistical inference from samples is based on the assumption that the sample is a random sample from some population.

The mathematical model that the statistician constructs for practical use in interpreting data is often discovered by noting what happens when sample after sample is drawn from the same population. All of us have made the observation, for example, that the mean varies from one sample to another. Since we assume that the method of selection is kept uniform through the sampling process, the discrepancies that we observe can logically be assigned only to the sampling process—the population remains constant. It seems reasonable, therefore, that we call these fluctuations from sample to sample, SAMPLING ERRORS. This fluctuation of the statistic from sample to sample can be shown to be random; hence, the statistic itself is entitled to its own frequency function which is often called the SAMPLING DISTRIBUTION. Most every application of statistical methods involves one or more sampling distributions.

Before we proceed further, let us mention that the quality control engineer utilizes most of the above theory when he constructs an $\overline{X}$ control chart. He thinks in terms of a measurable quantity (random variable) which has a stable pattern of variation (frequency function). When he takes samples (random samples), he calculates the sample mean (statistic). If he has assumed that the pattern of variation of the measure is a normal frequency function, then it can be shown by mathematics that the distribution of the sample mean (sampling distribution) is also normal. If the engineer does not assume the pattern of variation to be normal, he takes advantage of one of the most practical and one of the most useful results from mathematical statistics known as the CENTRAL LIMIT THEOREM.[2] The theorem states that with increasing sample size the distribution of means for samples from any population which has a mean and variance tends toward the normal distribution as

a limiting form. This is one of the reasons that we prefer to use the $\bar{X}$ chart rather than an individual's chart, especially when we are not too well acquainted with the parent population of the process.

CONTROL CHARTS AND SIGNIFICANCE TESTS

Now consider a few more control chart concepts which are already the property of the quality control engineer and which can be directly translated into a discussion of significance tests. When using the control chart technique, the quality control engineer has a sequence of samples generated by a manufacturing or measuring process. If every set of sample values behave like a random sample from a fixed population, the system is said to be in STATISTICAL CONTROL. The assumption on which most statistical methods are based, including significance tests, is that a state of statistical control exists. Lack of control implies a shift in the population. What is the logic behind such a statement?

Consider a familiar application of an $\bar{X}$ control chart. It is to be used "to provide a basis for current decisions during production as to when to hunt for causes of variation and take action intended to correct them." [1] Assume that the process has been in statistical control for a long enough time to determine the process average and the 3-sigma control limits. These control limits serve as the criterion to decide whether the process is still operating at the same long run process level. By using the control limits which were determined from past data to make our present decision, we are actually making the assumption that the long run process level has not changed and we are testing this assumption. That is, we are making an assumption concerning the nature of the frequency function of the observable random variable. This assumption is called a STATISTICAL HYPOTHESIS.

An assumption that completely specifies the population is called a SIMPLE HYPOTHESIS. If the assumption does not completely specify the population, the hypothesis is called a COMPOSITE HYPOTHESIS.

The rule which decides between the two courses of action—namely, that the process is "in control" or "out of control"—and is based on some statistic computed from a sample of a given size is called a STATISTICAL TEST. The point plotted is the SAMPLE POINT. To conclude that the process is "out of control" is to REJECT the hypothesis; and to conclude that the process is "in control" is to ACCEPT the hypothesis. We might very well speak of the region outside of the control limits as the CRITICAL REGION or REJECTION REGION, and the region within the control limits as the ACCEPTANCE REGION. It is clear that the outcome of the test is reduced to either accepting or rejecting the HYPOTHESIS UNDER TEST. We are considering a TWO DECISION problem.

In the above discussion of the $\bar{X}$ control chart, if the control limits are established from a statistical point of view, the decisions of acceptance or rejection are made on a probability basis. If a sample point falls in the critical region, there are two possibilities: either a very rare event has occurred, and the hypothesis is true, or there has been a SIGNIFICANT change in the process level and the hypothesis is false. We bet on the latter. Just how rare the event must be in order that the hypothesis be rejected must be defined. If the 3-sigma limits are used on the $\bar{X}$ control chart and if the sampling distribution of the means is strictly a normal frequency distribution, a process

481

which is unchanged will show sample points outside the 3-sigma limits
only 26 times in 10,000. Since we are taking sample values to make an
inference about the long run process level, we accept the fact that we
are going to reject a true hypothesis with a certain probability. That
is, we are willing to run the risk of making a wrong decision, for
example, 26 times in 10,000 and say that a sample point beyond the
3-sigma limits signifies a change in the process level. This risk that
we take in rejecting a true hypothesis (that is, of saying the process
level has changed when actually it has not) is called the LEVEL OF
SIGNIFICANCE and is denoted by the Greek letter α (alpha). A true
hypothesis will be rejected α percent of the time. Statisticians
call this error a TYPE I error.

The quality control engineer is also running the risk of making
another type of error—the error of stating that the process level has
not changed when in reality it has. The error of accepting the hypoth-
esis when it is false is labeled the TYPE II error. The risk (proba-
bility) of a TYPE II error is designated by the Greek letter β (beta).
This is to indicate that β percent of the time the test will allow an
actual change in process level (of preassigned size) to slip by un-
noticed. The technique of imposing the limit β at the desired point
will be illustrated numerically later.

We have stated that the decision to accept or reject a hypothesis
is based upon sample information and is characterized by risks that we
are willing to take that our decision be wrong. What we mean by a
wrong decision is conveniently illustrated in Figure 1. This figure
occurs in Wald's fundamental paper on the mathematical theory of
decision problems, and illustrates very clearly the four possible
situations resulting from a decision.[3]

TRUE SITUATION	HYPOTHESIS TRUE (Level Unchanged)	HYPOTHESIS FALSE (Level Changed)
ACTION TAKEN	DESCRIPTION OF DECISION	
ACCEPT the Hypothesis	O.K.	TYPE II ERROR
REJECT the Hypothesis	TYPE I ERROR	O.K.

Figure 1.   Four Possible Situations Resulting
from an Application of an Industrial
Significance Test.

Just as in the control chart procedure, let us consider that the reason for gathering and analyzing data is the need for analytical assistance in deciding between two courses of action, say A and B. The course of action that we should take would be clear if we only knew the precise nature of the population. However, because we lack this complete knowledge, we will make a decision based on some statistic of a random sample taken from the population.

Consider the following scheme to place the simple hypotheses into two separate piles and label them as follows: (2)

> H_0: all simple hypotheses whose truth would make A the better action.

> H_1: all simple hypotheses whose truth would make B the better action.

The following example serves to illustrate the above.

> Example 1. A coin which we have not examined is to be tossed. We assume that the coin is a "standard honest" coin. However, it might have two heads or two tails. We adopt the following test: toss the coin four times, if all heads or all tails occur, we reject the coin as standard. Let p denote the probability of tossing heads in a single toss. The hypothesis being tested, H_0, and the alternative, H_1, are:
>
> $$H_0: p = \tfrac{1}{2} \quad \text{and} \quad H_1: p = 0 \text{ or } p = 1$$
>
> Here, H_0 is a simple hypothesis and H_1 is a composite hypothesis. The probability of rejecting H_0 when H_0 is true is $\alpha = 1/8$. The probability of failing to detect a non-standard coin is $\beta = 0$ for either of the simple hypotheses which make up the composite H_1.

Tests of Population Mean – Population Normal and Standard Deviation Known – A Two-sided Test.

Consider the case in which the sample size is prescribed by considerations beyond our control. In this situation, we are able to determine only the significance level α, and then to accept whatever happens to us, good or bad, in the way of power with which to detect the fact that the hypothesis is false – although this may be our primary motive. The following steps will prove helpful for testing a statistical hypothesis.

1. Choose the random variable and decide on the a priori conditions which are not under test but are considered known.

2. State the hypothesis H_0 together with the alternative hypothesis H_1.

483

3. Specify the test statistic.

4. Denote the sample size and the significance level α.

5. Determine the critical region.

6. Take a random sample, calculate the value of the test statistic, and make a decision to either accept H_0 or to reject H_0.

Consider the following example of testing a statistical hypothesis.

> Example 2. We are interested in a measurement on a manufactured part, and would specify that the long run process average of the dimension should be, say, $\mu = \mu_0 = 25$ in order that the bulk of the parts function satisfactorily when in an assembly. We take a sample of 16 and find the sample mean to be $\bar{X} = 21$. Assume from past data that the distribution of the dimension may be considered normally distributed (for practical purposes) and that the standard deviation of the process seems to remain constant and is $\sigma = 6$. On the basis of the sample information, have we reason to believe that the process has shifted?

The following are the above recommended steps for testing a statistical hypothesis as they apply to example 2.

1. The random variable is the measurement on the manufactured part. The a priori conditions not under test are that the random variable is distributed normally with constant standard deviation $\sigma = 6$.

2. The hypotheses to be tested are

$$H_0: \mu = 25 \qquad \text{and} \qquad H_1: \mu \neq 25.$$

3. The test statistic, for this particular test, is

$$z = \frac{\bar{X} - \mu_0}{\sigma / \sqrt{n}}$$

The random variable z is distributed normally with mean 0 and standard deviation 1.

4. Let the significance level $\alpha = 5\%$ and the sample size $n = 16$.

5. We wish the critical region to reflect that a dimension either too large or too small will not function satisfactorily. The critical region expressed in terms of z becomes

$$z \leqslant -1.96 \qquad \text{and} \qquad z \geqslant 1.96.$$

(In terms of the sample average $\bar{X}$, the critical region is

$$\bar{X} \leqslant 22.06 \qquad \text{and} \qquad \bar{X} \geqslant 27.94.)$$

6. For our particular sample we calculate z to be

$$z = \frac{21 - 25}{6 / \sqrt{16}} = -2.67$$

Since this value of z lies in the critical region, the
hypothesis H_0 is rejected.

We have discussed the significance test with regard to the proba-
bility of rejecting the hypothesis when it is true that the mean is
actually 25. What is the probability of rejecting the hypothesis when
the process has actually shifted to 22? 26? It begins to look as
though we had considered only one point, μ_0 = 25, whereas all of the
values around 25 might be of great interest.

OPERATING CHARACTERISTIC CURVES AND POWER CURVES

We shall introduce the properties of the operating characteristic
curves for tests of significance by discussing the OC curves for accept-
ance sampling plans.

When considering a sampling plan, we are interested not only in the
AQL and its associated α ; in addition, we are interested in knowing
what the sampling plan will do when lots of quality other than the AQL
are submitted for acceptance. The chance that a lot has of being
accepted depends upon its lot quality. What ability has the plan to
detect these differences in lot quality? The OPERATING CHARACTERISTIC
curve gives this information, that is, the probability of accepting a
given lot of any percent defective.

Many of the ideas of accepting sampling, including the operating
characteristic curves, carry over immediately to significance tests.
How will our significance test behave when the true process level μ
is not μ_0 ? It is essential that we have an OC curve which will apply
to the significance test. These curves exist. [4] The OPERATING
CHARACTERISTIC CURVE FOR A STATISTICAL TEST gives the probabilities of
accepting the hypothesis corresponding to any alternative process level.
Some statisticians like to talk in terms of a POWER CURVE for the test.
The power curve gives the probabilities of rejecting the hypothesis
corresponding to any alternative process level. Since the probability
of rejection is one minus the probability of accepting, we see that the
power curve gives exactly the same information as the OC curve. Either
curve can be used to judge the adequacy of the test. However, in this
paper, we will construct the OC curves for the tests.

Operating Characteristic Curve for the Two-sided Test.

To derive the OC curve for the above test, we compute the proba-
bility that z falls between the following two quantities,

$$\text{OC function} = P\left[\frac{22.06 - \mu}{\sigma/\sqrt{n}} < z < \frac{27.94 - \mu}{\sigma/\sqrt{n}}\right]$$

where μ and σ are the parameters of the normal distribution and n
is the sample size. The quantities $\sigma = 6$ and n = 16 are given as
known.

For the various values of μ the OC can be developed as follows.

(1) μ	(2) $\dfrac{22.06 - \mu}{6/\sqrt{16}}$	(3) $\dfrac{27.94 - \mu}{6/\sqrt{16}}$	(4) OC Curve - Probability of Accepting H_0	(5) Power - Probability of Rejecting H_0
19	+ 2.04	+ 5.96	.021	.979
21	+ .71	+ 4.63	.240	.760
23	- 0.63	+ 3.29	.734	.266
25	- 1.96	+ 1.96	.950	.050
27	- 3.29	+ 0.63	.734	.266
29	- 4.63	- 0.71	.240	.760
31	- 5.96	- 2.04	.021	.979

Figure 2 gives the OC curve associated with the above tabular values.

Figure 2. OC Curve for the Two-sided Test of the
Hypothesis H_0: μ = 25 with α = .05
and n = 16 given σ = 6.

Long Run Process Average μ .
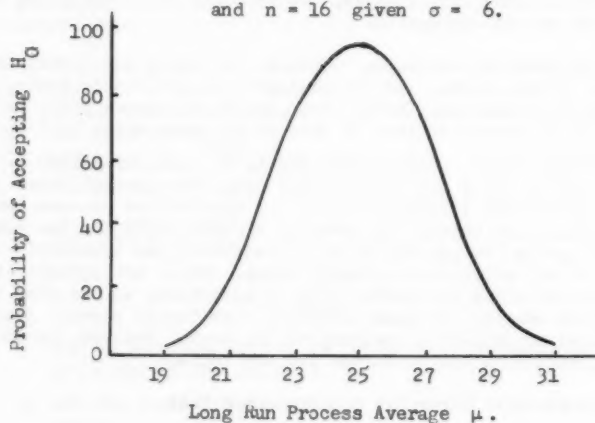
Tests of Population Mean - Population Normal and Standard Deviation
Known - A One-sided Test.

In the previous section we considered a two-sided test because of
the nature of the critical region which was dictated by the hypotheses
H_0 and H_1 .

Example 3. Consider a change in the nature of the hypotheses in example 2, and that

$$H_0: \mu \leqslant 25 \qquad \text{and} \qquad H_1: \mu > 25 .$$

The critical region defining the test becomes one-sided and of the form $z \geqslant k$, where the constant k is determined by the significance level α. If $\alpha = 0.05$, the critical region is given by $z \geqslant 1.645$. The other steps of the analysis remain the same as before. To derive the OC curve for this test we compute the following probability

$$OC \text{ function} = P \left[z < \frac{27.47 - \mu}{\sigma / \sqrt{n}} \right]$$

where μ and σ are the parameters of a normal distribution and n is the sample size. In our problem it is given a priori that $\sigma = 6$ and $n = 16$. The OC curve for this test is given in Figure 3.

Figure 3. Operating Characteristic for a
One-sided Test.

$$H_0: \mu \leqslant 25; \quad H_1: \mu > 25; \quad \alpha = 0.05; \text{ and } n = 16.$$

Long Run Process Average μ .
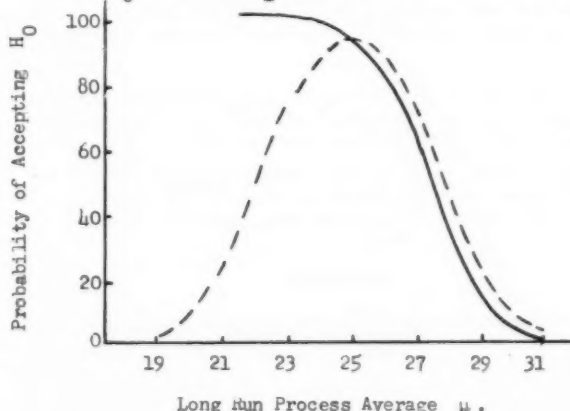
Determination of the Sample Size - Given Two Values of the Population Mean with Corresponding α and β.

In the above problems we arbitrarily chose the size of the sample and the significance level. However, step number 4 of the "steps for testing a statistical hypothesis," could read "Choose two values of the parameter and the corresponding values of α and β, and find the sample size n." The remainder of the steps do not change.

Example 4. Consider again example 2 which involves the two-sided test with the hypotheses

$$H_0: \mu = 25 \quad \text{and} \quad H_1: \mu \neq 25$$

where the standard deviation is known and $\sigma = 6$. However, in this situation suppose that we have objectively evaluated the probabilities of making wrong decisions and have set a level of significance $\alpha = 0.05$ for $\mu = 25$. In addition, we state that a process average which has shifted to 21 or 29 should only be accepted 10% of the time.

Let n be the size of the sample to be taken and let $\overline{X}_U$ and $\overline{X}_L$ be the upper and lower limits of the two sides of the critical region to be specified in terms of the sample average. We will assume that the sample averages are normally distributed. The following equations describe the above conditions which define the test:

$$\frac{\overline{X}_U - 29}{6/\sqrt{n}} = -1.282$$

$$\frac{\overline{X}_L - 21}{6/\sqrt{n}} = +1.282$$

$$\frac{\overline{X}_U - 25}{6/\sqrt{n}} = +1.960$$

$$\frac{\overline{X}_L - 25}{6/\sqrt{n}} = -1.960$$

Here we have three independent equations in three unknowns. If we eliminate $\overline{X}_U$ from the first and third equations, we obtain

$$n = \frac{6(1.960 + 1.282)}{4} = 23.65$$

$$n = 23.65$$

We will use $n = 24$. Now if we substitute $n = 24$ in both the first and second equations, we find that $\overline{X}_U = 27.4$ and that $\overline{X}_L = 22.6$ which give the boundaries of the upper and lower segments of the critical region that defines the test.

The operating characteristic curve for the above test can be found in a similar manner to that previously used for the two-sided test given for example 2.

Tests of Proportions - Large Samples

A similar procedure to that described above can be used to test hypotheses regarding population proportions. That is, we might be interested in testing whether an industrial process is operating in a random manner to produce a specified fraction defective of units. It is very common to determine the critical region and sample size corresponding to a given set of hypotheses by using the normal approximation to the binomial distribution.

SUMMARY

In many instances the quality control engineer is faced with a situation which demands that he decide on either of two causes of action, A or B. There are statistical methods which deal with the gathering and analyzing of data for the purpose of analytical assistance in deciding on the course of action to be taken. The suggested steps offered to test significance can serve to help the engineer state the problem together with deciding upon the probabilities of making wrong decisions.

We should emphasize that though we draw a conclusion as the result of the statistical test, the test doesn't prove anything. If it turns out to accept H_0, this does not mean that H_0 is necessarily true. We do not have to believe either that it is or that it is not true; we have a decision as to which of two actions to take. The test is simply a rule for decision, and one which turns out all right in many cases.

REFERENCES

1. Grant, Eugene L., "Statistical Quality Control," McGraw-Hill Co., Inc., New York, 1946.

2. Lindgren & McElrath, "Brief Course in Statistical Theory," Burgess Publishing Co., Minneapolis, Minn., 1957.

3. Wald, Abraham, "On the Principles of Statistical Inference," University of Notre Dame, Indiana, 1942.

4. Ferris, C. D., Grubbs, F.E., and Weaver, C. L., "Operating Characteristics for the Common Statistical Tests of Significance," Annals of Mathematical Statistics, Vol. 17, (1946).

5. Dixon, W. J. and Massey, F. J., "Introduction to Statistical Analysis," McGraw-Hill Co., Inc., New York, 1951.

6. Duncan, A. J., "Quality Control and Industrial Statistics," Richard D. Irwin, Inc., Homewood, Ill., 1952.

490

SIGNIFICANCE TESTS II - SMALL SAMPLES

Max Astrachan
Air Force Institute of Technology

In the first paper in this series we have discussed the basic philosophy and concepts behind tests of significance, and how such tests are related to control charts. In addition, we have seen how to construct one-sided and two-sided tests on the mean in the case in which the population standard deviation σ is known. Our procedure was based on the fact that the means of samples of any size from a normal population are normally distributed with mean equal to the population mean, and variance (square of standard deviation) equal to the population variance divided by N, the sample size. We used the test statistic

$$z = \frac{\bar{X} - \mu_o}{\frac{\sigma}{\sqrt{N}}} , \tag{1}$$

which is normally distributed with mean 0 and variance 1. $\bar{X}$ is the observed sample mean and μ_o the hypothesized population mean.

We cannot use z for a test statistic when σ is unknown, although it seems logical to replace σ in (1) by s, the sample standard deviation*, the resulting expression

$$t = \frac{\bar{X} - \mu_o}{\frac{s}{\sqrt{N}}} \tag{2}$$

is no longer normally distributed. Instead it follows what is known as "Student's" t distribution.

---

*Sample standard deviation is defined as the square root of the sum of the squares of the deviations of the measurements from their mean divided by N-1. In symbols,

$$s = \sqrt{\frac{\sum_{i=1}^{N} (x_i - \bar{x})^2}{N-1}}$$

The quantity N-1 is called "degrees of freedom," abbreviated df, and we commonly say that s^2 has df $= N - 1$. For ease of calculation, especially when a desk calculator is available, it is often convenient to replace the numerator by the following expression, to which it is algebraically equivalent:

$$\sum_{i=1}^{N} x_i^2 - \frac{\left(\sum_{i=1}^{N} x_i\right)^2}{N} .$$

491

For a given sample size and for a population of known mean, z contains but one quantity, $\bar{X}$, which changes from sample to sample. However, t has two such quantities, $\bar{X}$ and s. Hence it seems reasonable that t fluctuates more from sample to sample than does z.

There are, in fact, many t-distributions, one for each value of df (degrees of freedom) associated with the sample variance. For (2) we always have df $= N - 1$. Each distribution curve is bell-shaped and looks like the normal distribution curve but has longer "tails." Figure 1 shows several t-curves and the normal curve. As N increases, the t-distribution approaches the normal distribution with mean 0 and variance 1.
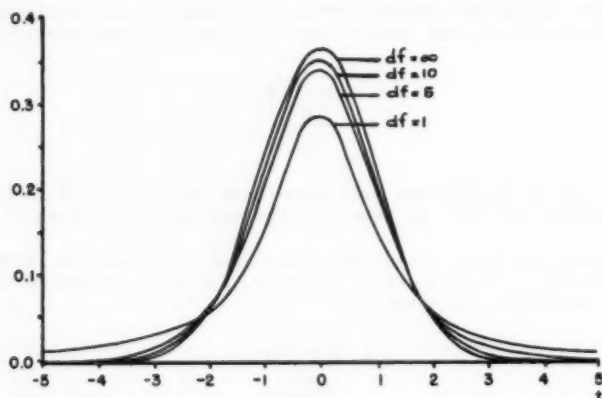
Fig. 1. Distribution curves of t for df = 1, 5, 10 and ∞.
(The curve for df = ∞ is the standard normal curve)

In order to use the t-statistic to test hypotheses, we shall need to know percentile points corresponding to the levels of significance we want to set. Tables of these values are given in most standard books on statistics. (See for example $\boxed{1}$, p 384.) Since there is a different curve for each sample size, there will be many sets of such points. Table 1 gives a portion of a table usually found in the literature. The particular t-distribution is specified in the column headed df ($= N - 1$),

Table 1 - Some Percentile Points of the t-Distribution

df	Percentile		
	.95	.975	.99
5	2.015	2.571	3.365
10	1.812	2.228	2.764
15	1.753	2.131	2.602
20	1.725	2.086	2.528
25	1.708	2.060	2.485
30	1.697	2.042	2.457
60	1.671	2.000	2.390
120	1.658	1.980	2.358
∞	1.645	1.960	2.326

and the percentiles are indicated at the top of the table. Because of the symmetry of the distribution, the .05, .025, and .01 percentile points are the negatives of the .95, .975, and .99 values respectively. Note that the values in the last row, for which df $= \infty$, are those for the normal distribution. Since they do not differ much from the percentile points for large df we often use z in (1) as a test statistic when σ is unknown, but the sample size is large. In such cases, we replace σ by its estimate, the sample standard deviation s.

With this brief description of the t-distribution, we are now in a position to perform tests of hypotheses on the mean in the case in which the population standard deviation σ is unknown. The procedures of the first paper in this series are applicable, the only difference will be in the test statistic and the critical values. As we shall see later, we shall be able to perform other tests of hypotheses also, by use of the t-distribution.

We shall illustrate the method by means of the following example:

Example 1. The weights in pounds of a sample of 6 castings are as follows: 49, 55, 53, 57, 48, 53. Is the process consistent with the specified mean of 50 lbs?

Applying the procedure of the first paper we then have the following steps:

1. The random variable is the weight of a casting. We assume these weights are normally distributed.

2. Hypothesis to be tested: The long run process average is 50 lbs., i.e., H_0: $\mu = 50$. The alternative hypothesis is H_1: $\mu \neq 50$.

3. The test statistic is

$$t = \frac{\overline{X} - \mu_o}{\frac{s}{\sqrt{N}}} = \frac{\overline{X} - 50}{\frac{s}{\sqrt{6}}}$$

which we assume has a t-distribution with 5 degrees of freedom (df $= N - 1$ in this case).

4. Choose a significance level, $\alpha = 0.05$.

5. From Table 1, we see that the critical region for this significance level is t ≤ -2.571 and t ≥ 2.571.

6. From the given data, $\Sigma X_1 = 315$ and $\Sigma X_1^2 = 16,597$, whence $\overline{X} = 52.5$, and from the formula in the first footnote,

$$s = \sqrt{\frac{16,597 - \frac{(315)^2}{6}}{5}} = \sqrt{\frac{59.5}{5}} = \sqrt{11.9} = 3.45.$$

493

Substituting into the expression for t in step 3 above, we have

$$t = \frac{52.5 - 50}{\frac{3.45}{\sqrt{6}}} = \frac{2.5}{1.41} = 1.77$$

Since this value of t is not in the critical region we accept the hypothesis H_o, that the process mean is 50 lbs.

The above is an example of a two-sided test, which was dictated by the manner in which the question was asked in the problem. Had the question been "Is the process consistent with the specification of mean weight not greater than 50 lbs?," our hypotheses would have been

$$H_o: \mu \leq 50, \quad H_1: \mu > 50.$$

The critical region then becomes one-sided and is of the form $t \geq k$, where k is determined by the level of significance and the df. In our case, where $\alpha = 0.05$ and $df = 5$, the critical region would be $t = 2.015$. The remainder of the analysis is the same. Whether we are going to make a one-sided or a two-sided test must, of course, be decided in advance.

COMPARING TWO POPULATION MEANS

The problem of comparing two populations with respect to some characteristic is one that frequently arises in practice. In these situations, one course of action would be appropriate if the two populations may be considered the same with respect to the characteristic in question, and another if they are not.

The following questions suggest the sort of problem one meets in this area: Does a population of smokers have greater susceptibility to disease of a certain type than a population of non-smokers? Does fertilizer "A" work better than fertilizer "B" on the average? Does a certain new piece of equipment produce bar stock with greater variability in dimensions than the old? Are men better drivers than women? Does one method of measuring the concentration of a chemical have more inherent variability than another? Does machine "A" produce more defective parts than machine "B"?

Two random variables may differ with respect to level or centering as measured, for example, by their means; they may also differ with respect to variability or repeatability as measured by their standard deviations. There are other differences, of course, but these two are usually the ones in which we are interested. In this section we shall describe several tests which may be used for comparing two population means. As we shall see, the appropriate one to use depends upon the information available. Tests on variability will be covered in the next paper.

494

Suppose we have two samples of size N_1 and N_2 from each of two populations with means μ_1, μ_2, and variances σ_1^2, σ_2^2, respectively. To test the hypothesis that the two population means are equal, we shall use the test statistic

$$z = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\dfrac{\sigma_1^2}{N_1} + \dfrac{\sigma_2^2}{N_2}}} \tag{3}$$

where $\overline{X}_1$ and $\overline{X}_2$ are two sample means. If the two populations are normal and $\mu_1 = \mu_2$, then z is a standard normal variable (with mean 0 and variance 1).

The following example illustrates the use of this test statistic:

Example 2. A shearing pin has a hexagonal head and is made on two different machines. A sample of 100 is taken from each of the machines and the Rockwell hardness determined for the head of each pin. The following results were obtained:

Machine	N	$\overline{X}$	s
1	100	41.63	1.8
2	100	40.67	2.0

Is there a significant difference between the average hardness of the heads produced by the two machines?

Although we do not know the population variances, we shall assume that the sample sizes are large enough so that we can replace them by the sample variances in (3).

The test then proceeds as follows:

1. There are two random variables: X_1 and X_2, the hardnesses of the pins from the two machines. If these are normally distributed, the distribution of z in (3) is normal with mean 0 and variance 1. If not, we assume the sample sizes are large enough so that z is approximately normally distributed, even when we replace the population variances by sample variances.

2. Hypothesis to be tested: H_0: $\mu_1 = \mu_2$. The alternative hypothesis is H_1: $\mu_1 \neq \mu_2$.

3. The test statistic is given by (3), which we assume to be normally distributed.

4. Choose the level of significance $\alpha = 0.05$.

5. The critical region is $z \leq -1.96$ and $z \geq 1.96$.

495

6. Substituting the given data in (3), using the known sample variances in place of the unknown population variances, we have

$$z = \frac{41.63 - 40.67}{\sqrt{\dfrac{(1.8)^2}{100} + \dfrac{(2.0)^2}{100}}} = \frac{.96}{.269} = 3.57.$$

Since this value of z is in the critical region we reject H_0 and conclude that on the average the hardnesses of the pin heads produced by the two machines are different.

When the samples are small we cannot replace the population variances in (3) by sample variances, for then the statistic is no longer normally distributed. We have to distinguish between two cases, one in which the variances of the two populations are equal, and the other in which they are unequal. Tests for the equality of variances will be given in the next paper in this series.

Case 1 - Equal Population Variances. In this case we combine the two sample variances to give us an estimate of the common population variance. This is given by

$$s_p^2 = \frac{(N_1 - 1)s_1^2 + (N_2 - 1)s_2^2}{N_1 + N_2 - 2}. \tag{4}$$

We refer to s_p^2 as the pooled or averaged estimate of the common population variance σ^2 obtained from the two sample variances s_1^2 and s_2^2. The test statistic used to determine whether two population means are equal or not is then

$$t = \frac{\overline{x}_1 - \overline{x}_2}{s_p \sqrt{\dfrac{1}{N_1} + \dfrac{1}{N_2}}}, \tag{5}$$

which follows the t-distribution with $N_1 + N_2 - 2$ degrees of freedom if $\mu_1 = \mu_2$.

Example 3. Following are samples of the breaking strengths of two types of cloth:

Type 1:	61	52	45	46			
Type 2:	48	71	69	68	58	70	64

Is there a significant difference between the mean breaking strengths?

From the given data

$$\overline{x}_1 = 51, \quad \overline{x}_2 = 64, \quad s_1^2 = 54, \quad s_2^2 = 69.67$$

496

whence from (4)

$$s_p^2 = \frac{3 \times 54 + 6 \times 69.67}{4 + 7 - 2}$$

$$= \frac{580}{9} = 64.44,$$

so that

$$s_p = \sqrt{64.44} = 8.03.$$

Following the procedure of the other examples, then, we have the following steps:

1. There are two random variables X_1 and X_2, the breaking strengths of the two types of cloth. We are assuming the variances of the two populations are equal.

2. We are testing the hypothesis $H_0: \mu_1 = \mu_2$, against $H_1: \mu_1 \neq \mu_2$.

3. The test statistic is given by (5) which we assume follows the t-distribution with 9 degrees of freedom.

4. Choose the level of significance $\alpha = 0.05$.

5. The critical region is $t \leq -2.262$ and $t \geq 2.262$. (This is found from a more complete table than Table 1.)

6. Substituting the values found above in (5), we have

$$t = \frac{51 - 64}{8.03\sqrt{\frac{1}{4} + \frac{1}{7}}} = \frac{-13}{(8.03)(.6268)}$$

$$= -2.58.$$

Since this value of t is in the critical region we reject H_0 and conclude that the mean breaking strengths of the two types of cloth are not the same.

Case 2. Population Variances Unequal. In this case, assuming the populations are normally distributed, the statistic

$$t = \frac{\overline{X}_1 - \overline{X}_2}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}} \tag{6}$$

has approximately a t-distribution if the two population means are equal. The number of degrees of freedom is given by

$$df = \frac{\left(\dfrac{s_1^2}{N_1} + \dfrac{s_2^2}{N_2} \right)^2}{\dfrac{\left(\dfrac{s_1^2}{N_1} \right)^2}{N_1 + 1} + \dfrac{\left(\dfrac{s_2^2}{N_2} \right)^2}{N_2 + 1}} - 2, \qquad (7)$$

which is usually not an integer. However, the closest value in the table will generally be sufficient.

Example 4. The diameters of ten samples each of wire produced on two different machines were measured with the following results (data in thousandths of an inch):

Machine 1:   59   56   68   59   56   62   59   58   57   56

Machine 2:   57   55   59   56   57   54   54   57   55   54

Is there a difference between the mean diameters of the product of the two machines?

From the given data,

$$\overline{X}_1 = 59.0, \quad \overline{X}_2 = 55.8, \quad s_1^2 = 13.56, \quad s_2^2 = 2.84,$$

and from (7)

$$df = \frac{(1.356 + .284)^2}{\dfrac{(1.356)^2}{11} + \dfrac{(.284)^2}{11}} - 2 = 15.4 - 2 = 13.4.$$

The steps in performing the test are then as follows:

1. There are two random variables X_1 and X_2, the thicknesses of the wire from the two machines. We are assuming the populations are normally distributed and their variances unequal.

2. The hypothesis to be tested is $H_0: \mu_1 = \mu_2$, against $H_1: \mu_1 \neq \mu_2$.

3. The test statistic is given by (6) which we assume has approximately a t-distribution with degrees of freedom given by (7).

4. Choose the level of significance $\alpha = 0.05$.

5. For 13 degrees of freedom the critical region is $t \leqq -2.160$ and $t \geqq 2.160$, whereas for 14 degrees of freedom, it is $t \leqq -2.145$ and $t \geqq 2.145$.

6. Substituting the values found above in (6), gives

$$t = \frac{59.0 - 55.8}{\sqrt{\dfrac{13.56}{10} + \dfrac{2.84}{10}}} = \frac{3.2}{1.28} = 2.50.$$

Hence, we reject the hypothesis of equal means.

Again, as in the case of Example 1, we could have conducted one-sided tests in the last three examples had the questions been suitably asked. Thus, in Example 2, suppose we had been interested in knowing whether or not the average hardness of the pins produced by machine 1 was significantly greater than that produced by machine 2. Then we would be testing the hypothesis

$$H_0: \mu_1 \leqq \mu_2, \text{ against } \mu_1 > \mu_2.$$

and, at the .05 level of significance, the critical region would have been $z \geqq 1.645$.

PAIRED MEASUREMENTS

We have been assuming in all of the tests comparing two population means that the samples are independent. A special case when these are not independent occurs when the observations are paired. For example, we might be comparing two different devices for measuring tensile strength in order to determine whether or not they give the same results. The values obtained would vary together from piece to piece, and, hence, would not be independent. In this case the test is carried out on the differences of the paired measurements assuming they form a population whose true difference is zero. The procedure is illustrated in the following example.

Example 5. Two analysts, "A" and "B", analyzed a series of chemical mixtures to obtain the percentage of one of the ingredients common to all of them. Following are the results:

Mixture	1	2	3	4	5	6	7	8
Analyst A:	5	7	6	8	6	9	7	6
Analyst B:	9	5	8	8	7	8	8	7
Difference, d_i:	-4	2	-2	0	-1	1	-1	-1

Are the analysts reading alike?

499

These are obviously not independent samples since each pair of readings was made on a different mixture. From the row of differences, we have then

$$\sum d_i = -6, \ \sum d_i^2 = 28; \ \ \bar{d} = -.75,$$

$$s_d = \sqrt{\frac{28 - \frac{(-6)^2}{8}}{7}} = \sqrt{3.36} = 1.83.$$

As in the case of the preceding examples, we have the following steps:

1. Our random variable is d, the difference between the readings of the two analysts on the mixtures.

2. Hypothesis to be tested is that the mean of these differences is zero, i.e., H_0: $\mu_d = 0$, against H_1: $\mu_d \neq 0$. (μ_d is the mean of the population of differences.)

3. The test statistic is

$$t = \frac{\bar{d} - \mu_d}{\frac{s_d}{\sqrt{N}}}$$

which we assume follows a t-distribution with df = N - 1 = 7. (Here N is the number of sample differences.)

4. Choose a significance level $\alpha = 0.05$.

5. The critical region is then $t \leq -2.365$ and $t \geq 2.365$.

6. Substituting the results found above into the test statistic of step 3, we have

$$t = \frac{-.75 - 0}{\frac{1.83}{\sqrt{8}}} = \frac{-.75}{.647} = -1.16$$

Since this value of t is not in the critical region, we assume the two analysts are reading alike.

We have not included the operating characteristic curves for the tests described in this paper. They are given in [2] and [3], together with some examples of their uses. Table 12 in [1] includes power curves for some of the tests. The form in which it is given makes the table particularly useful for some types of problems.

500

Other methods of testing hypotheses of the form discussed in this paper are also available. Some of these involve use of the range instead of the standard deviation. Several alternate tests are described in $\underline{/4/}$ and $\underline{/5/}$. The quality control engineer who is planning to use significance tests would do well to examine other ways of making them in order to determine which are best for his particular purpose.

It should be emphasized again that significance tests are simply guides to help us make up our minds as to which course of action seems best to follow in a given situation, other things being equal. The quality control engineer must also consider the practical implications of the recommendations he makes to management. For example, a statistical significance test may show that a new machine would increase our production rate by 10 units per day. However, the cost of buying and installing the new machine may be more than the added profit realized from the greater production. Significance tests prove nothing in themselves, but they can be helpful in deciding what to do when we are faced with alternatives.

<div align="center">References</div>

1. Introduction to Statistical Analysis, W. J. Dixon and F. J. Massey, Jr. McGraw - Hill Book Company, Inc., 1957.

2. "Operating Characteristics for the Common Statistical Tests of Significance," by Charles D. Ferris, Frank E. Grubbs, and Chalmers L. Weaver. Annals of Mathematical Statistics, June 1946.

3. Handbook of Industrial Statistics, A. H. Bowker and G. J. Lieberman. Prentice Hall, Inc., 1955.

4. "Rough-and-Ready Statistical Tests," by W. Allen Wallis. Industrial Quality Control, March 1952.

5. "Some Notes on the Theory and Application of Rank Order Statistics," by Ralph A. Bradley. Industrial Quality Control, February 1955.

SIGNIFICANCE TESTS III - VARIANCES

Fred C. Leone
Case Institute of Technology

To continue the discussion of significance tests, we now turn to tests on the variance. We have considered the basic thinking of tests of significance, the risks, the assumptions, and procedures. Further tests on the sample mean, $\bar{X}$, were explained for both large and small samples. One of the points mentioned earlier in tests of significance on the mean of two samples was an assumption of equality of variances. That is, in a particular test of significance on sample mean $\bar{X}_A$ versus sample mean $\bar{X}_B$, it was assumed that the variances σ_A^2 and σ_B^2 were equal. This is equivalent to saying that the variances of the population from which each sample came were the same. We now proceed to take this assumption and treat it as a hypothesis which we test.

Tests of hypotheses on the variance may be divided into two groups, namely, (a) sample versus population, and (b) sample versus sample. In the former the type of question we consider is whether this sample we are proposing has come from a normal distribution with a variance equal to some specified value. In the latter case we consider a question like the following: Could these two samples have come from normal distributions with the same variance? (Here we do not specify what this population variance is.) We may also include in this latter category tests on the homogeneity of several variances as well as tests on one of a set of variances being significantly larger or significantly smaller than the others.

Consider the following example. Castings have been produced with a variance (σ^2) equal to four pounds. A new process has been proposed which is much more economical. This change will be adopted provided that the variance of this is the same as the proposed variance of four pounds. A sample is to be taken and a decision is to be made as a result of the test.

Let us restate the procedure which was proposed in the introductory paper of this series. This is as follows:

1. Choose the random variable and decide on the a priori conditions which are not under test but are considered known.

2. State the hypothesis, H_0, together with the alternative hypothesis, H_1.

3. Specify the test statistic.

4. Denote the sample size and the significance level.

5. Determine the critical region.

6. Take a random sample, calculate the value of the test statistic, and make a decision to either accept H_0 or to reject H_0.

The steps of this test for the example, then, are as follows:

1. The random variable is the weight of the casting. The a priori assumptions are (a) that this variable is distributed normally with some unknown mean (μ) and unknown variance (σ^2) and (b) the sample is randomly chosen and its elements are independent.

2. The hypothesis and its alternative are

$$H_0 : \sigma^2 = \sigma_o^{\ 2} \ (= 4)$$
$$H_1 : \sigma^2 \neq \sigma_o^{\ 2}$$

3. The test statistic is

$$\chi^2 = \frac{(N - 1)s^2}{\sigma_o^{\ 2}}$$

where s^2 is the variance from the sample and is defined as

$$s^2 = \frac{\sum_{i=1}^{N} (X_i - \overline{X})^2}{N - 1}$$

4. Let us specify a sample of size 6 and the significance level α equal to 0.05.

5. The critical region is

$$\chi^2 \leq 0.831 \quad \text{and} \quad \chi^2 \geq 12.83$$

6. The sample, the test, the conclusion: If $0.831 < \chi^2 < 12.83$, accept the null hypothesis (H_0). If $\chi^2 \leq 0.831$ or $\chi^2 \geq 12.83$, reject the hypothesis.

   In order to determine the critical region of rejection of the test it is necessary to make use of Table 1. This table presents the percentiles of the test statistic, χ^2. The χ^2 distribution is a family of probability curves. The members of the family change with different values of degrees of freedom, ν. Here ν is N-1. Figure 1 presents several of the curves of the probability distribution. Note how the curves tend toward a symmetrical distribution as N increases. The abscissa of these curves is χ^2 and the ordinate is $f(\chi^2)$. The area under each curve is, of course, equal to 1. Incidentally, the mean of each of these curves is ν.

   Turning now to Table 1, the row designations are the degrees of freedom and the column headings, the percentages. The numbers within the body of the table are the values of χ^2 which give the designated probability to their left. For example, for ν equals 5 degrees of freedom, the value $\chi^2 = 0.831$ has 2.5 percent of the area under the curve to

504
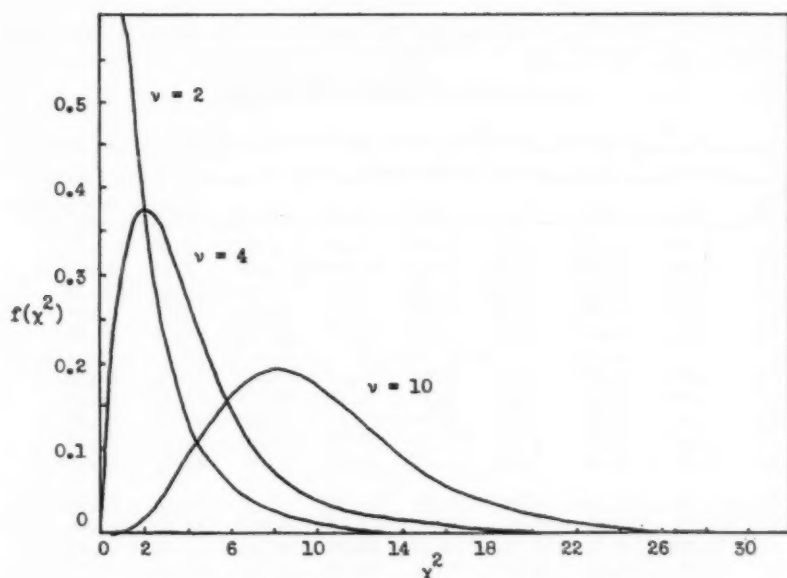
Figure 1. χ^2 Distribution for $\nu = 2$, 4, and 10.

the left of it. The value $\chi^2 = 12.83$ has 97.5 percent of the area to the left. (This is equivalent to 2.5 percent to the right.) If we wish our α to be 0.05, we can have 2.5 percent on each tail of the distribution.

Let us continue the problem. A random sample is taken with the following weights of castings in pounds, as presented in a problem in the previous paper:

$$49, 55, 53, 57, 48, 53$$

From this we calculate $s^2 = 11.9$. Finally

$$\chi^2 = \frac{(N-1)s^2}{\sigma_o^2}$$

$$= \frac{5(11.9)}{4} = 14.9$$

χ^2 is then in the critical region, that is,

$$\chi^2 \ (= 14.9) > 12.83$$

505

TABLE 1

PERCENTILES OF THE χ^2 DISTRIBUTION

ν \ P	Per Cent							
	0.5	1.0	2.5	5.0	95.0	97.5	99.0	99.5
1	0.000	0.000	0.001	0.004	3.84	5.02	6.63	7.88
2	0.010	0.020	0.051	0.103	5.99	7.38	9.21	10.60
3	0.072	0.115	0.216	0.352	7.81	9.35	11.34	12.84
4	0.207	0.297	0.484	0.711	9.49	11.14	13.28	14.86
5	0.412	0.554	0.831	1.145	11.07	12.83	15.09	16.75
6	0.676	0.872	1.24	1.64	12.59	14.45	16.81	18.55
7	0.989	1.24	1.69	2.17	14.07	16.01	18.48	20.28
8	1.34	1.65	2.18	2.73	15.51	17.53	20.09	21.96
9	1.73	2.09	2.70	3.33	16.92	19.02	21.67	23.59
10	2.16	2.56	3.25	3.94	18.31	20.48	23.21	25.19
15	4.60	5.23	6.26	7.26	25.00	27.49	30.58	32.80
20	7.43	8.26	9.59	10.85	31.41	34.17	37.57	40.00
30	13.79	14.95	16.79	18.49	43.77	46.98	50.89	53.67
40	20.71	22.16	24.43	26.51	55.76	59.34	63.69	66.77
60	35.53	37.48	40.48	43.19	79.08	83.30	88.38	91.95

We, therefore reject the null hypothesis. On the basis of the sample presented we have information to state that the new process, though more economical, is not producing a product which is equally as variable as the established process. Our sample points to a significant difference in variances.

Let us stop to compare the meaning of rejection as opposed to acceptance. If we reject a hypothesis with a specified α, then we can state that we have a 1-α probability that this sample did not come from the proposed population. On the other hand, if a hypothesis is accepted, we simply state that we are willing to "go along" with it. We shall act as if it is true. In some instances we may have to state that we simply do not have enough information to negate the hypothesis.

Sometime ago this analogy was presented to me: A result of a test of hypothesis is equivalent to the result of a law court. A man is either found guilty (rejection of hypothesis of innocence) or not guilty (acceptance of the hypothesis). In the former, we have enough information to convict him. In the latter, we either do not have enough information, if he is guilty, or he is really innocent. We do not know which is correct.

At this point it is well to consider the operating characteristic (OC) curve of this test of significance. Recall that α is the probability of rejecting the null hypothesis, H_0, when it is true. β is the probability of accepting the alternative hypothesis H_1 when H_0 is

true. For a sample of size 6 the critical values of χ^2 are

$$\chi^2 = 0.831 \text{ and } 12.83$$

Our first step is to translate these critical values of χ^2 to critical values of s^2. By using the equation below, with s^2 unknown, we have for each χ^2

$$\frac{(N-1)s^2}{\sigma_o^2} = \chi^2$$

Substituting for N, σ_o^2 and the critical value of χ^2 we have

$$\frac{5s_1^2}{4} = 0.831$$

and

$$\frac{5s_2^2}{4} = 12.83$$

This gives us the critical values of s^2, namely,

$$s_1^2 = .665$$

$$s_2^2 = 10.26$$

Suppose we now consider alternative hypotheses. For the hypothesis $H_1 : \sigma^2 = 9 \ (= \sigma_1^2)$ the value of β (error of the second kind) can be calculated as follows:

$$\chi_1^2 = \frac{5(.665)}{9} = .369$$

$$\chi_2^2 = \frac{5(10.26)}{9} = 5.700$$

β is the probability that χ^2 is between χ_1^2 and χ_2^2 for 5 degrees of freedom, that is,

$$P(.369 < \chi^2 < 5.700) = .65$$

This same procedure is continued for other alternative hypotheses. These results are presented in Table 2. The first column presents σ_1

507

TABLE 2

CALCULATION OF OPERATING CHARACTERISTIC CURVE FOR $H_0 : \sigma^2 = \sigma_0^{\,2}$ $(= 4)$
FOR 5 DEGREES OF FREEDOM

σ_1	$\sigma_1^{\,2}$	$x_1^{\,2} = \dfrac{5(.665)}{\sigma_1^{\,2}}$	$x_2^{\,2} = \dfrac{5(10.26)}{\sigma_1^{\,2}}$	$\beta = P(x_1^{\,2} < x^2 < x_1^{\,2})$
(1)	(2)	(3)	(4)	(5)
1	1	3.325	51.30	.65
2	4	0.831	12.83	.95
3	9	0.369	5.700	.65
4	16	0.208	3.206	.33
5	25	0.133	2.052	.16
6	36	0.092	1.425	.08
7	49	0.068	1.047	.04
8	64	0.052	0.802	.02
9	81	0.041	0.633	.01

for the alternative hypotheses $(\sigma^2 = \sigma_1^{\,2})$, where $\sigma_1^{\,2}$ is given in column 2. Column 3 is the lower limit of x^2 for the alternative hypothesis, while column 4 gives the upper limit. Finally, β is calculated in column 5. This is the area under the x^2 curve for 5 degrees of freedom between the two limits, $x_1^{\,2}$ and $x_2^{\,2}$.

The results of Table 2 are presented in Figure 2. The abscissa is the value of $\sigma_1^{\,2}$ for the alternative hypothesis. A scale for σ_1 is also stated. The ordinate is the error of the second kind (β) if the alternative hypothesis is true and H_0 is accepted. For comparison, a second operating characteristic curve is presented by a broken curve for the similar problem with 15 degrees of freedom rather than 6. Many authors prefer to use $1-\beta$ as the ordinate. This figure of the compliments of β is called the power of the test.

In the previous example you will note that the test was two sided. We asked beforehand whether σ^2 was significantly different than $\sigma_0^{\,2}$. Suppose, on the other hand, we ask whether σ^2 is significantly greater than $\sigma_0^{\,2}$. This is a one sided test, and our single critical value for the same sample size and $\alpha = 0.05$ would be $x^2 = 11.07$. If then

$$x^2 < 11.07$$

we accept the null hypothesis. Otherwise we reject the hypothesis.
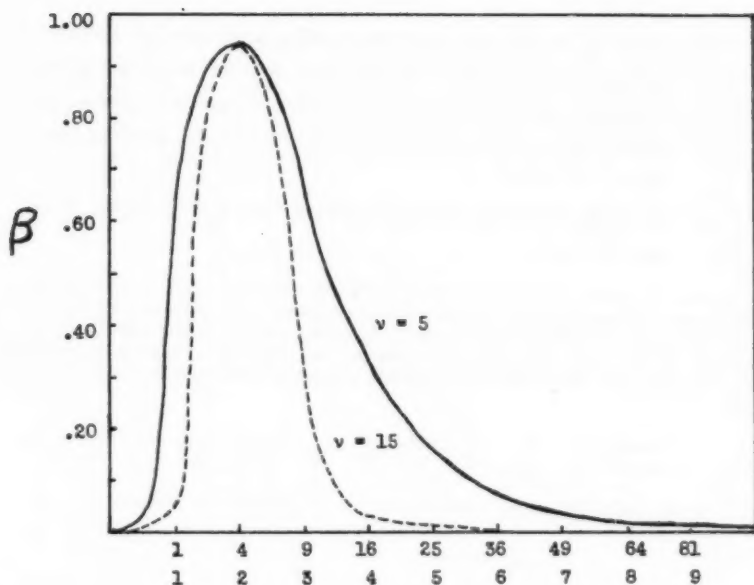
Figure 2. Operating Characteristic Curves for $H_0 : \sigma^2 = \sigma_o^2 \; (= 4)$ for 5 and 15 degrees of freedom.

Let us turn now to the problem of comparing sample variances from two different populations. From each population we take a sample and determine its variance. Are these variances compatible? In other words, could these samples have come from normal distributions with equal variances?

In setting up the general steps in testing a statistical hypothesis we have:

1. The random variable is the measurement, X_i, and the a priori conditions are again, normality, randomness, and independence of the values in each of the two samples.

2. The hypothesis is

$$H_0 : \sigma_1^2 = \sigma_2^2$$

where σ_1^2 is the variance of the population from which the sample one came, and σ_2^2 is the other variance.

3. The test statistic is

$$F = s_1^2 \, / \, s_2^2$$

509

where s_1^2 is the variance from sample one, and s_2^2 is the variance from sample two. For this test we shall designate as sample one that sample which has a larger s^2. Hence, we require $s_1^2 > s_2^2$. The purpose of this is to simplify the use of the table.

4. The sample sizes are respectively N_1 and N_2. Consider an α, say, $\alpha = 0.05$.

Let us again present an example which was proposed in the preceding paper. The breaking strength of two types of cloth was tested for a significant difference of means. The variances of the sample were assumed equal. We now take this assumption and test it as a hypothesis. The two samples of breaking strength in pounds per square inch (psi) are:

Sample A: 61 52 45 46
Sample B: 48 71 69 68 58 70 64

If we calculate the variances from the sample we obtain

$$s_A^2 = 54$$

$$s_B^2 = 69.7$$

Let us call $s_B^2 = s_1^2$, and $s_A^2 = s_2^2$ in order that $s_1^2 > s_2^2$.

The particular steps in the test of significance then become:

1. X is the breaking strength in psi.

2. $H_0 : \sigma_1^2 = \sigma_2^2$
   $H_1 : \sigma_1^2 \neq \sigma_2^2$

3. $F = s_1^2 / s_2^2$

4. Sample (1) of size 7 (Sample B)
   Sample (2) of size 4 (Sample A)

5. The critical region is $F \geq F_{.975, 6, 3}$

6. The samples are taken and F is calculated. If $F < 14.7$, accept the hypothesis that the sample variances are equal. If $F \geq 14.7$, reject the hypothesis.

To determine the critical values it is necessary to consult Table 3. In this we have the percentiles of the F distribution for ν_1 degrees of freedom in the numerator, and ν_2 degrees of freedom in the denominator. The F distribution is a family of curves which has

two parameters, v_1 and v_2. The statistic F is the ratio of two sample variances where each of the samples was originally drawn from a normal distribution. In the table each cell of four values is made up of the values of F with v_1 and v_2 degrees of freedom, which give the probability as stated in each row. This probability is the area to the left under the F distribution.

For a two sided test of hypothesis, the critical value of

$$F = s_1^2 / s_2^2 > 1$$

for a specified α is that tabular value which has a probability of $1-\alpha/2$. The nature of F is such that by requiring that F be greater than 1 we need only consider an $\alpha/2$ area on the upper tail. In Table 3, then, for $v_1 = 6$ and $v_2 = 3$, the .975 row gives us a value of 14.7.

To conclude this example, we calculate

$$F = \frac{69.7}{54}$$

$$= 1.29$$

Since $F < F_{.975, 6, 3}$ (= 14.7), we accept the hypothesis of equal variance. As done in the preceding paper, we can then continue the test for equality of means.

Note that the above was a two sided test. If, however, before actually taking the samples, we suspected that one variance would be larger and asked, for example, "Is variance B greater than variance A?" we would have a one sided test. The critical value of F would then be $F_{.95, 6, 3}$ which is equal to 8.94.

There is a further set of tests which should be mentioned at this time. These concern the homogeneity of several variances. Suppose k samples were taken and we wish to know if the variances were compatible. The hypothesis would then be:

$$H_0 : \sigma_1^2 = \sigma_2^2 = \ldots = \sigma_k^2$$

One of the tests commonly used is Bartlett's test. This can be found in most books on statistical methods.

A test known as Cochran's test was devised to consider whether a particular one of the variance of the k samples is significantly larger than the others. This test is especially useful prior to the analysis of variance when it is suspected that a particular cell in the design does not have a variance compatible with the others. This equality of variances is often assumed in the analysis of variance technique.

Recently a test was devised by R. Doornbos[*] to test the significance of the smallest of a group of variances.

In conclusion, let us state that we have considered a comparison of a sample variance with a proposed population variance. The proposed test was supplemented by an operating characteristic curve. Recall, also, that if a hypothesis is rejected, we have a certain assurance of our action being correct. We had decided on a risk of rejection of the null hypothesis if it were true. On the other hand, if the hypothesis is accepted, we do not have this high assurance that it is correct. We simply go along with it. We can, however, compare the likelihood of this hypothesis with an alternative hypothesis by considering the operating characteristic curve. This same procedure of comparing a sample variance with a population variance was carried into a comparison of two sample variances. Finally, it must be kept in mind that to simply test a hypothesis and agree or disagree with it is often not enough. We must continue our investigation and consider confidence intervals on our population parameter. If we reject a hypothesis, for example, the logical question to follow with is, "What about the true parameter of the population from which the sample did come? What can we say about that?" Here we get into the area of confidence intervals which is to be discussed in the next paper.

[*]"Significance of a smallest of a set of estimated normal variances," by R. Doornbos, _Statistics Neerlandica_, 10, No. 2, 117-26, 1956.

TABLE 3

Percentiles of the $F(v_1, v_2)$ Distribution with v_1 Degrees of Freedom for the Numerator and v_2 for the Denominator.

v_2	P	1	2	3	4	5	6	7	8	9	10	15	20	30	40	60	∞
1	.95	161	200	216	225	230	234	237	239	241	242	246	248	250	251	252	254
	.975	648	800	864	900	922	937	948	957	963	969	985	993	100^1	101^1	101^1	102^1
	.99	$*405^1$	500^1	540^1	562^1	576^1	586^1	593^1	598^1	602^1	606^1	616^1	621^1	626^1	629^1	631^1	637^1
	.995	$*162^2$	200^2	216^2	225^2	231^2	234^2	237^2	239^2	241^2	242^2	246^2	248^2	250^2	251^2	253^2	255^2
2	.95	18.5	19.0	19.2	19.2	19.3	19.3	19.4	19.4	19.4	19.4	19.4	19.4	19.5	19.5	19.5	19.5
	.975	38.5	39.0	39.2	39.2	39.3	39.3	39.4	39.4	39.4	39.4	39.4	39.4	39.5	39.5	39.5	39.5
	.99	98.5	99.0	99.2	99.2	99.3	99.3	99.4	99.4	99.4	99.4	99.4	99.4	99.5	99.5	99.5	99.5
	.995	198	199	199	199	199	199	199	199	199	199	199	199	199	199	199	200
3	.95	10.1	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.70	8.66	8.62	8.59	8.57	8.53
	.975	17.4	16.0	15.4	15.1	14.9	14.7	14.6	14.5	14.5	14.4	14.3	14.2	14.1	14.0	14.0	13.9
	.99	34.1	30.8	29.5	28.7	28.2	27.9	27.7	27.5	27.3	27.2	26.9	26.7	26.5	26.4	26.3	26.1
	.995	55.6	49.8	47.5	46.2	45.4	44.8	44.4	44.1	43.9	43.7	43.1	42.8	42.5	42.3	42.1	41.8
4	.95	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.86	5.80	5.75	5.72	5.69	5.63
	.975	12.2	10.6	9.98	9.60	9.36	9.20	9.07	8.98	8.90	8.84	8.66	8.56	8.46	8.41	8.36	8.26
	.99	21.2	18.0	16.7	16.0	15.5	15.2	15.0	14.8	14.7	14.5	14.2	14.0	13.8	13.7	13.7	13.5
	.995	31.3	26.3	24.3	23.2	22.5	22.0	21.6	21.4	21.1	21.0	20.4	20.2	19.9	19.8	19.6	19.3
5	.95	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.62	4.56	4.50	4.46	4.43	4.36
	.975	10.0	8.43	7.76	7.39	7.15	6.98	6.85	6.76	6.68	6.62	6.43	6.33	6.23	6.18	6.12	6.02
	.99	16.3	13.3	12.1	11.4	11.0	10.7	10.5	10.3	10.2	10.1	9.72	9.55	9.38	9.29	9.20	9.02
	.995	22.8	18.3	16.5	15.6	14.9	14.5	14.2	14.0	13.8	13.6	13.1	12.9	12.7	12.5	12.4	12.1

*See last page of Table for explanation.

TABLE 5 (Continued)

Percentiles of the $F(v_1, v_2)$ Distribution

v_2	v_1 / P	1	2	3	4	5	6	7	8	9	10	15	20	30	40	60	∞
6	.95	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	3.94	3.87	3.81	3.77	3.74	3.67
	.975	8.81	7.26	6.60	6.23	5.99	5.82	5.70	5.60	5.52	5.46	5.27	5.17	5.07	5.01	4.96	4.85
	.99	15.7	10.9	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.56	7.40	7.23	7.14	7.06	6.88
	.995	18.6	14.5	12.9	12.0	11.5	11.1	10.8	10.6	10.4	10.2	9.81	9.59	9.36	9.24	9.12	8.88
7	.95	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.51	3.44	3.38	3.34	3.30	3.23
	.975	8.07	6.54	5.89	5.52	5.29	5.12	4.99	4.90	4.82	4.76	4.57	4.47	4.36	4.31	4.25	4.14
	.99	12.2	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.31	6.16	5.99	5.91	5.82	5.65
	.995	16.2	12.4	10.9	10.0	9.52	9.16	8.89	8.68	8.51	8.38	7.97	7.75	7.53	7.42	7.31	7.08
8	.95	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.22	3.15	3.08	3.04	3.01	2.93
	.975	7.57	6.06	5.42	5.05	4.82	4.65	4.53	4.43	4.36	4.30	4.10	4.00	3.89	3.84	3.78	3.67
	.99	11.3	8.65	7.59	7.01	6.63	6.37	6.18	6.05	5.91	5.81	5.52	5.36	5.20	5.12	5.05	4.86
	.995	14.7	11.0	9.60	8.81	8.30	7.95	7.69	7.50	7.34	7.21	6.81	6.61	6.40	6.29	6.18	5.95
9	.95	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.01	2.94	2.86	2.83	2.79	2.71
	.975	7.21	5.71	5.08	4.72	4.48	4.32	4.20	4.10	4.05	3.96	3.77	3.67	3.56	3.51	3.45	3.33
	.99	10.6	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	4.96	4.81	4.65	4.57	4.48	4.31
	.995	13.6	10.1	8.72	7.96	7.47	7.13	6.88	6.69	6.54	6.42	6.03	5.83	5.62	5.52	5.41	5.19
10	.95	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.85	2.77	2.70	2.66	2.62	2.54
	.975	6.94	5.46	4.83	4.47	4.24	4.07	3.95	3.85	3.78	3.72	3.52	3.42	3.31	3.26	3.20	3.08
	.99	10.0	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.56	4.41	4.25	4.17	4.08	3.91
	.995	12.8	9.43	8.08	7.34	6.87	6.54	6.30	6.12	5.97	5.85	5.47	5.27	5.07	4.97	4.86	4.64
15	.95	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.40	2.33	2.25	2.20	2.16	2.07
	.975	6.20	4.76	4.15	3.80	3.58	3.41	3.29	3.20	3.12	3.06	2.86	2.76	2.64	2.59	2.52	2.40
	.99	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.52	3.37	3.21	3.13	3.05	2.87
	.995	10.8	7.70	6.48	5.80	5.37	5.07	4.85	4.67	4.54	4.42	4.07	3.88	3.69	3.59	3.48	3.26

TABLE 8 (Continued)

Percentiles of the $F(v_1, v_2)$ Distribution

v_2	P	1	2	3	4	5	6	7	8	9	10	15	20	30	40	60	∞
20	.95	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.20	2.12	2.04	1.99	1.95	1.84
	.975	5.87	4.46	3.86	3.51	3.29	3.13	3.01	2.91	2.84	2.77	2.57	2.46	2.35	2.29	2.22	2.09
	.99	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.09	2.94	2.78	2.69	2.61	2.42
	.995	9.94	6.99	5.82	5.17	4.76	4.47	4.26	4.09	3.96	3.85	3.50	3.32	3.12	3.02	2.92	2.69
30	.95	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.01	1.93	1.84	1.79	1.74	1.62
	.975	5.57	4.18	3.59	3.25	3.03	2.87	2.75	2.65	2.57	2.51	2.31	2.20	2.07	2.01	1.94	1.79
	.99	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.70	2.55	2.39	2.30	2.21	2.01
	.995	9.18	6.35	5.24	4.62	4.23	3.95	3.74	3.58	3.45	3.34	3.01	2.82	2.65	2.52	2.42	2.18
40	.95	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	1.92	1.84	1.74	1.69	1.64	1.51
	.975	5.42	4.05	3.46	3.13	2.90	2.74	2.62	2.53	2.45	2.39	2.18	2.07	1.94	1.88	1.80	1.64
	.99	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.52	2.37	2.20	2.11	2.02	1.80
	.995	8.83	6.07	4.98	4.37	3.99	3.71	3.51	3.35	3.22	3.12	2.78	2.60	2.40	2.30	2.18	1.93
60	.95	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.84	1.75	1.65	1.59	1.53	1.39
	.975	5.29	3.93	3.34	3.01	2.79	2.63	2.51	2.41	2.33	2.27	2.06	1.94	1.82	1.74	1.67	1.48
	.99	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.35	2.20	2.03	1.94	1.84	1.60
	.995	8.49	5.80	4.73	4.14	3.76	3.49	3.29	3.13	3.01	2.90	2.57	2.39	2.19	2.08	1.96	1.69
∞	.95	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.67	1.57	1.46	1.39	1.32	1.00
	.975	5.02	3.69	3.12	2.79	2.57	2.41	2.29	2.19	2.11	2.05	1.83	1.71	1.57	1.48	1.39	1.00
	.99	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.04	1.88	1.70	1.59	1.47	1.00
	.995	7.88	5.30	4.28	3.72	3.35	3.09	2.90	2.74	2.62	2.52	2.19	2.00	1.79	1.67	1.55	1.00

1. The number 1 as a superscript indicates a product by 10
2. The number 2 as a superscript indicates a product by 100

Read 405[1] as 4050, and 162[2] as 16200.

SIGNIFICANCE TESTS IV: CONFIDENCE INTERVALS

Robert W. Boeke
John Deere Ottumwa Works

In the preceding papers we have been primarily concerned with decision statistics. We have stated a hypothesis, selected a level of significance, calculated a critical region, then tested the hypothesis with the statistic calculated from our observations; based on the location of our statistic relative to the critical region we have either accepted or rejected our hypothesis. Our observations either substantiate or refute the hypothesis under test and we decide either to accept or reject it. It should be observed that we have utilized the statistics to test our hypothesis rather than to estimate the population parameters. Frequently our sampling may be performed to estimate a population parameter.

If sampling is performed to estimate a population parameter, we may use either a point or interval estimate. For instance, we may draw a random sample from a population and use the sample average as a point estimate of the population average. Thus, X is a point estimate of μ. However, since X is subject to sampling fluctuation we may, particularly for small samples, obtain a value which is misleading. The interval estimate is of such a nature that for a population parameter, Θ, limits may be calculated (L_1 and L_2) so that the interval may be expected to encompass the parameter with a probability of $1-\alpha$, where α is the probability that the interval does not include the parameter. In generalized form the interval may be written:

$$L_1 < \Theta < L_2$$

and the confidence coefficient is $1-\alpha$.

If a random sample of N items is drawn from a normal population we may estimate an interval which has a probability of, for instance, 95% of encompassing the population mean.

$$L_1 < \mu < L_2$$

and 95% of such intervals would be expected to include μ. The values L_1 and L_2 are known as confidence limits and these limits define the confidence interval.

We cannot say that μ will be in the interval calculated but we may say that the probability is .95, before the sample is drawn, that the interval so defined will include μ. A rather fine distinction must be drawn regarding the statements which can be made. Since the parameter has a fixed value, it does not have a sampling distribution. Therefore, our probability statement must be made concerning the calculated interval, not the parameter. For instance, the statement, "The probability is .95 that μ lies within the interval", is incorrect since it infers that μ is a variable. We may, however, state that we are 95% confident that our calculated interval encompasses the parameter. Or, if we repeatedly estimate the parameter in this manner, we may expect to include the parameter in our intervals 95% of the time. After the interval is calculated the parameter is either included or it is not; therefore, the probability that the parameter lies within the interval is either 0 or 1.

Table I illustrates the variable characteristics of the confidence intervals. Fifteen samples were drawn from a population with $\mu = 40$ and $\sigma = 10$. For each sample, 95% confidence intervals were calculated (1) assuming σ known and (2) assuming σ not known. It should be observed that for this example μ is fixed and equal to 40, the confidence interval is the variable. If this experiment were repeated indefinitely, approximately 95% of the intervals calculated would encompass the mean, $\mu = 40$.

Perhaps the biggest advantage of an interval estimate is that we not only get a point estimate of the parameter but also obtain a measure of the precision of our estimate. We may increase the precision, i.e. decrease the size of the interval, by increasing the sample size. Two samples may have the same average but the one with the smaller confidence interval indicates greater precision of estimation. A comparison of Tables I and II will give an idea of the relative precision for sample sizes of 5 and 25 respectively. Both tables were computed using 95% confidence intervals; the narrower limits in Table II reflect the increased precision obtained by the larger sample sizes.

Throughout the discussion, reference has been made to 95% confidence intervals. The confidence coefficient need not be 95% but may be selected by the experimentor based upon the degree of confidence required under the specific conditions. Although values of 90, 95, and 99% are frequently employed, the individual should review his particular problem and need not feel hesitant to deviate from these values.

A. Confidence Intervals for μ : σ known

The sampling distribution of $\overline{X}$ has a mean equal to μ, a standard deviation, $\sigma/\sqrt{N}$, and is distributed according to the normal curve. The values of $\overline{X}$ are grouped about μ in such a manner that the interval $\mu \pm 1.96$ $\sigma/\sqrt{N}$ will include some known percentage, 95%, of $\overline{X}$ - values. Then the following interval will include the difference, $\mu - X$, 95 percent of the time

$$- 1.96\,\sigma/\sqrt{N} < \mu - \overline{x} < + 1.96\,\sigma/\sqrt{N}$$

By adding $\overline{X}$ to each term, we have

$$\overline{X} - 1.96\,\sigma/\sqrt{N} < \mu < \overline{X} + 1.96\,\sigma/\sqrt{N}$$

In the general case we have

$$\overline{X} + Z_{\frac{1}{2}\alpha}\,\sigma/\sqrt{N} < \mu < \overline{X} + Z_{1-\frac{1}{2}\alpha}\,\sigma/\sqrt{N}$$

where $Z_{\frac{1}{2}\alpha}$ and $Z_{1-\frac{1}{2}\alpha}$ are obtained from Table IV, and α is the risK we are willing to take that the interval may not include μ .

Example I.
    We wish to estimate the average tensile strength of a new run of twine. Previous tests indicate substantial variation in the mean tensile strength, μ, but indicate that σ is constant at 16 pounds. We wish to estimate the 90 percent confidence interval based upon a sample of 25 observations with an average of 375 pounds.

518

Given:

$$\sigma = 16 \text{ (known)}$$
$$\alpha = .10 \text{ or } 1 - \alpha = .90$$
$$N = 20$$
$$\bar{X} = 375$$

Then,

$$375 - \frac{(1.645)(16)}{\sqrt{25}} < \mu < 375 + \frac{(1.645)(16)}{\sqrt{25}}$$

$$369.7 < \mu < 380.3$$

We are 90 per cent confident that the interval, 369.7 to 380.3 lbs., will include the mean breaking strength of the twine. Our point estimate, or "best" estimate is 375 pounds but our interval indicates the precision inherent in our estimate.

It should be observed that the confidence interval may also be used in conjunction with a test of significance. Had we tested the hypothesis that $\mu = 370$ with $\alpha = .10$, we would not have sufficient reason to doubt that $\mu = 370$. Our critical region would include values less than 369.7 or greater than 380.3; since 370 lies within the confidence limits we accept the hypothesis that $\mu = \mu_0 = 370$.

It must be emphasized that results are defendable for small sample sizes only if the population is normally distributed. For large sample sizes the distribution of averages approaches normality regardless of the form of the parent population. To obtain tenable results, an adequate sample size hinges upon the amount by which the parent population deviates from the normal distribution and the variation in the data.

B. Confidence Intervals for μ : σ not known

Where the population standard deviation, σ , is known we have used the sampling distribution of $z = \bar{X} - \mu /\sigma/\sqrt{N}$ which is normally distributed. However, we seldom know σ . Therefore, we use a test statistic $\bar{X} - \mu / s/\sqrt{N}$ which follows student's t - distribution. Although the t - distribution approaches the normal distribution for large sample sizes, it must compensate for the sampling fluctuation in both $\bar{X}$ and s which is significant for smaller samples. Since we want the confidence intervals to include μ a given percentage of the time, intuitively we acknowledge that for comparable conditions (i.e. sample sizes and risks) the confidence intervals tend to be longer where σ is unknown than where σ is known.

Reference to Tables I and II will reveal that where σ is known the confidence intervals are of constant length. However, with σ unknown both the length and location vary and the average length of the interval is longer. The longer length of the interval does not mean that our estimate of μ is any less precise; however, it reflects the lack of confidence in estimating μ since we are relying on a sample estimate, s, rather than the parameter, σ .

Where σ is unknown, the confidence interval for μ is

$$\bar{X} + t_{\frac{1}{2}\alpha} \, s/\sqrt{N} < \mu < \bar{X} + t_{1-\frac{1}{2}\alpha} \, s/\sqrt{N}$$

519

The values for t may be read from Table V for the values of $\propto$ selected and N-1 degrees of freedom.

Example II.

A new supplier of twine submits samples for tensile tests. We wish to estimate the population mean, μ, with 90% confidence. Thirty samples are tested with $\bar{X} = 368$ and $s = 28$.

$$368 - \frac{(1.70)(28)}{\sqrt{30}} < \mu < 368 + \frac{(1.70)(28)}{\sqrt{30}}$$

$$368 - 8.7 < \mu < 368 + 8.7$$

$$359.3 < \mu < 376.7$$

Our best estimate of the tensile strength of the new twine is our sample average, 368, and we are 90% confident that the interval 359.3 to 376.7 actually encompasses the population mean.

Here, again, we may observe the similarity between the confidence limits and the critical region used in tests of significance. We should also emphasize that this procedure is valid only where the population has a normal distribution.

C. Confidence Intervals for $\mu_1 - \mu_2$: $\sigma_1 = \sigma_2 = \sigma$ is known

If we wish to estimate the difference between two population means, μ_2, and μ_2, where the population variances are known and equal $\sigma_1^2 = \sigma_2^2 = \sigma^2$, the confidence interval is

$$\bar{X}_1 - \bar{X}_2 + Z_{\frac{1}{2}\propto} \sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} < \mu_1 - \mu_2 < \bar{X}_1 - \bar{X}_2 + Z_{1-\frac{1}{2}\propto} \sigma \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

Example III.

Samples of twine are taken from two different runs and tested for tensile strength. Experience indicates that the means may be different but the standard deviation is known to be 16. We wish to estimate the difference between the means, μ_1, and μ_2, of the two runs. The following information is given

$$N_1 = 25$$
$$N_2 = 16$$
$$\bar{X}_1 = 373$$
$$\bar{X}_2 = 351$$
$$s_1 = 14$$
$$s_2 = 17$$

If we wish to be 99% confident that our limits encompass, $\mu_1 - \mu_2$, then $Z_{\frac{1}{2}\propto} = -2.576$ and $Z_{1-\frac{1}{2}\propto} = +2.576$ (Table IV.)

Then
$$373-351 - (2.576)(16) \sqrt{\frac{1}{25} + \frac{1}{16}} < \mu_1 - \mu_2 < 373-351 + (2.576)(16) \sqrt{\frac{1}{25} + \frac{1}{16}}$$
$$22 - 13.1 < \mu_1 - \mu_2 < 22 + 13.1$$
$$8.9 < \mu_1 - \mu_2 < 35.1$$

520

Our best estimate of the difference in strength between the runs is 22. Moreover we are 99% confident that this true average is between 8.9 and 35.1.

By comparison with a test of significance based on the hypothesis that $\mu_1 = \mu_2$, we would have rejected the hypothesis since 0 does not lie within the confidence interval. As in Section A, the validity of the test depends upon the normal distribution of the populations for small sample sizes.

D. Confidence Intervals for $\mu_1 - \mu_2$: $\sigma_1 = \sigma_2 = \sigma$ is unknown

Where σ is not known, the test statistic must rely upon the t-distribution rather than the normal distribution. For estimating the difference between two means where the standard deviations are assumed equal, but unknown, the confidence limits are:

$$\bar{x}_1 - \bar{x}_2 + t_{\frac{1}{2}\alpha} \, s_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}} < \mu_1 - \mu_2 < \bar{x}_1 - \bar{x}_2 + t_{1-\frac{1}{2}\alpha} \, s_p \sqrt{\frac{1}{N_1} + \frac{1}{N_2}}$$

where $t_{\frac{1}{2}\alpha}$ and $t_{1-\frac{1}{2}\alpha}$ are determined for $N_1 + N_2 - 2$ degrees of freedom s_p is the pooled standard deviation and is equal to:

$$s_p = \sqrt{\frac{(N_1 - 1) s_1^2 + (N_2 - 1) s_2^2}{N_1 + N_2 - 2}}$$

Example IV.

Assume the same information as is given in Example III but assume that the standard deviation, σ, is unknown.

$$s_p = \sqrt{\frac{(24)(196) + (15)(289)}{25 + 16 - 2}}$$

$$= 15.2$$

Then,

$$373 - 351 - (2.70)(15.2)\sqrt{\frac{1}{25} + \frac{1}{16}} < \mu_1 - \mu_2 < 373 - 351 + (2.70)(15.2)\sqrt{\frac{1}{25} + \frac{1}{16}}$$

$$22 - 13.2 < \mu_1 - \mu_2 < 22 + 13.2$$

$$8.8 < \mu_1 - \mu_2 < 35.2$$

The results are comparable to those of Example III. However, it should be observed that the reduced estimate of the standard deviation (15.2 as compared with 16.0) is more than offset by the increase in $t_{\frac{1}{2}\alpha}$ over $z_{\frac{1}{2}\alpha}$.

E. Confidence Intervals for Paired Comparisons

If tests are conducted so that for each condition or level of testing, observations are obtained from each of two different populations, then paired and the differences computed, we may use the following method of calculating confidence intervals:

$$\bar{x}_d + t_{\frac{1}{2}\alpha} \, s_d / \sqrt{N} < \mu_d < \bar{x}_d + t_{1-\frac{1}{2}\alpha} \, s_d / \sqrt{N}$$

521

where $\bar{X}_d$ is the average of the differences, s_d is the standard deviation of the differences and N is the number of differences, not total observations. Also $t_{\frac{1}{2}\alpha}$ and $t_{1-\frac{1}{2}\alpha}$ are obtained for N-1 degrees of freedom.

Example V.
We wish to estimate the average difference in tensile strength for two brands of twine so that we may be 95% confident of including the true difference.

Two brands of twine are tested for tensile strength under various weather conditions. In each instance a test is made of each brand. Tests are made under five different conditions with the following results:

Test No.	Brand A	Brand B	Difference
1	376	349	+27
2	362	342	+20
3	320	305	+15
4	345	322	+23
5	384	369	+15

Then $\bar{X}_d = 20$, $s_d = 5.2$, $N = 5$, $t_{\frac{1}{2}\alpha} = -2.78$ and $t_{1-\frac{1}{2}\alpha} = +2.78$

$$20 - (2.78)(5.2) \big/ \sqrt{5} < \mu_d < 20 + (2.78)(5.2) \big/ \sqrt{5}$$

$$20 - 6.5 < \mu_d < 20 + 6.5$$

$$13.5 < \mu_d < 26.5$$

We are 95% confident that the limits, 13.5 and 26.5 encompass the true difference.

As in sections B and D the t- distribution is restricted to cases where the population is approximately normal. Where we pair observations, this limitation will be fulfilled if each population has a normal distribution and the effects, under various conditions, are additive.

F. Confidence Intervals for σ^2

The confidence limits for σ^2 are based upon the $X^2/d.f.$ distribution. The ratio s^2/σ^2 is distributed according to $X^2/d.f.$. If we let $P_{\frac{1}{2}\alpha}$ and $P_{1-\frac{1}{2}\alpha}$ represent the values for $X^2/d.f.$ for various degrees of freedom, then,

$$P_{\frac{1}{2}\alpha} < s^2/\sigma^2 < P_{1-\frac{1}{2}\alpha}$$

By inverting and multiplying by s^2, we have

$$\frac{s^2}{P_{\frac{1}{2}\alpha}} > \sigma^2 > \frac{s^2}{P_{1-\frac{1}{2}\alpha}}$$

which are the confidence limits for σ^2 .

Example VI.
In example II we investigated a new brand of twine and estimated its tensile strength. Using the same data ($\bar{X} = 368$, $s = 28$, and $N = 30$)

522

let us calculate the 90% confidence limits for σ^2 .

$$\frac{784}{.610} > \sigma^2 > \frac{784}{1.45}$$

$$1286 > \sigma^2 > 540$$

We are 90% confident that the limits 540 and 1286 encompass the true value of σ^2 .

By taking the square root of each term we obtain confidence limits for σ .

$$34.2 > \sigma > 23.2$$

The confidence interval for σ is interpreted the same as that for σ^2 but may have the advantage of being somewhat more understandable.

When using the $X^2/d.f.$ distribution we should be confident that our population is normally distributed.

Table I

95% Confidence Intervals for μ
    with N = 5 drawn from a population with μ = 40 and
        σ^2 = 100

(1) σ^2 known = 100

(2) σ^2 Unknown

524
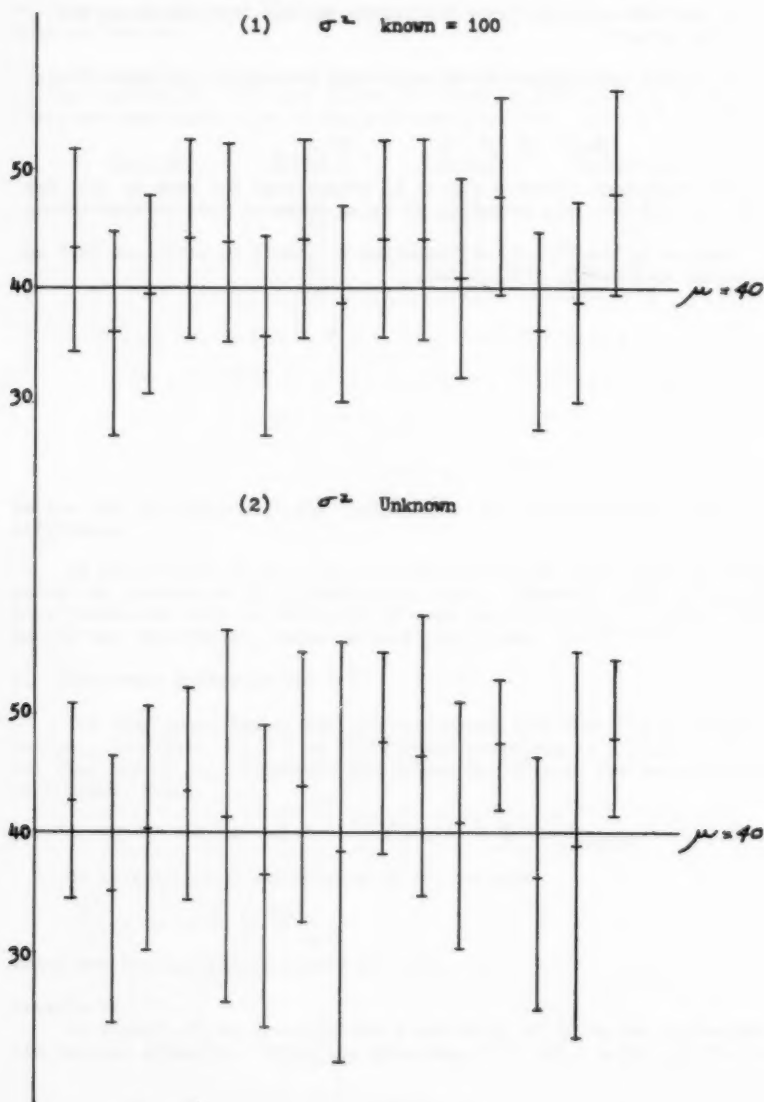
Table II

Confidence Interval For μ with N = 25 drawn from a
population with μ = 40 and σ^2 = 100

(1) σ^2 - known = 100

(2) σ^2 - unknown

525
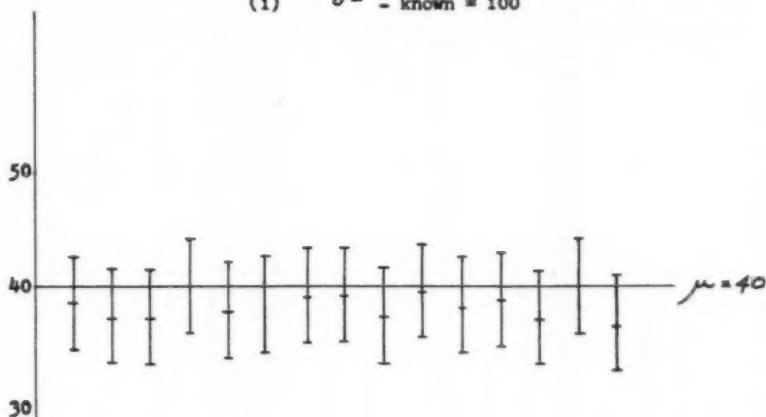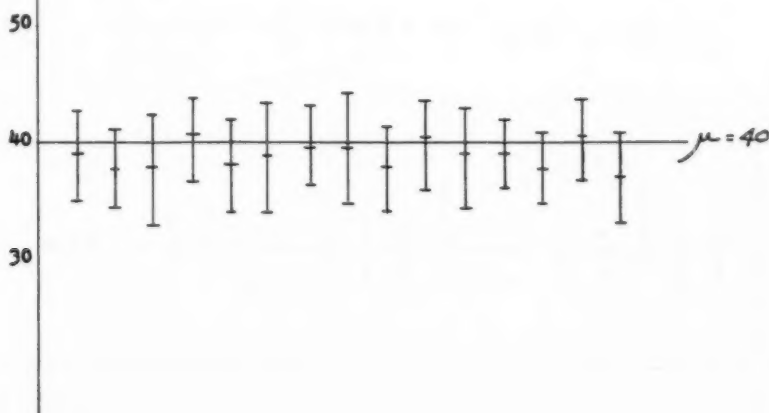
Table III

95 Confidence Intervals for σ
  Drawn from a population with μ = 40 and σ^2 = 100

(1) N = 5
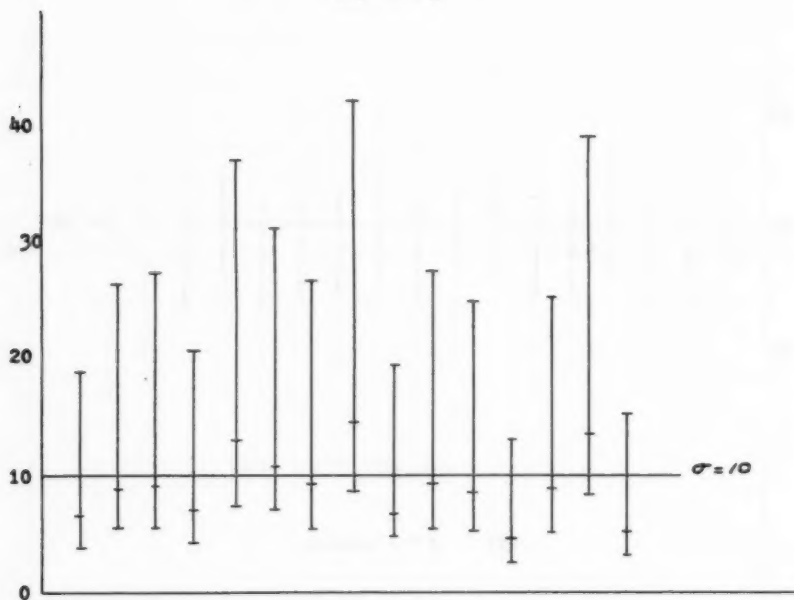
(2) N = 25
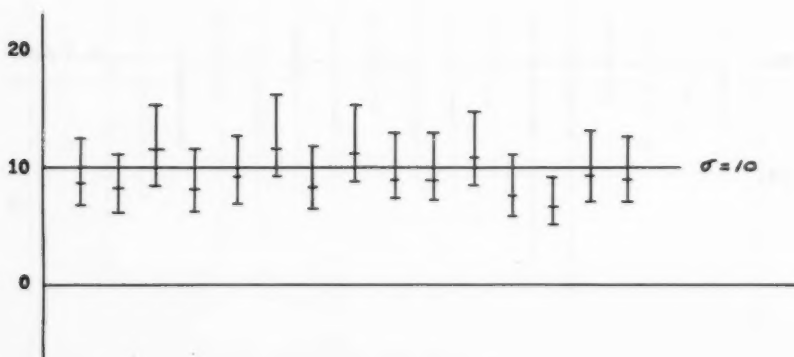
Table IV Areas Under The Normal Curve
-∞ To Z

Z	AREA
-3.090	.001
-2.576	.005
-2.326	.010
-1.960	.025
-1.645	.050
-1.282	.100
- .842	.200
- .253	.400
0	.500
+ .253	.600
+ .842	.800
+1.282	.900
+1.645	.950
+1.960	.975
+2.326	.990
+2.576	.995
+3.090	.999

Table V. Values Of t

Degrees of Freedom	t .95	t .975	t .995
1	6.31	12.7	63.7
2	2.92	4.30	9.92
3	2.35	3.18	5.84
4	2.13	2.78	4.60
5	2.01	2.57	4.03
6	1.94	2.45	3.71
7	1.89	2.36	3.50
8	1.86	2.31	3.36
9	1.85	2.26	3.25
10	1.81	2.23	3.17
12	1.78	2.18	3.05
14	1.76	2.14	2.98
16	1.75	2.12	2.92
18	1.73	2.10	2.88
20	1.72	2.09	2.85
30	1.70	2.04	2.75
40	1.68	2.02	2.70
60	1.67	2.00	2.66
120	1.66	1.98	2.62
∞	1.645	1.960	2.576
d.f.	$t_{.05}$	$t_{.025}$	$t_{.005}$

Prefix the t - values with a minus sign when reading the table from the bottom.

Table VI. Values of $X^2/d.f.$

Degrees of Freedom	P .005	P .025	P .050	P .950	P .975	P .995
1	.00004	.001	.004	3.84	5.02	7.88
2	.005	.025	.051	3.00	3.69	5.30
3	.024	.072	.117	2.60	3.12	4.28
4	.052	.121	.178	2.37	2.79	3.72
5	.082	.166	.229	2.21	2.57	3.37
6	.113	.206	.273	2.10	2.41	3.09
7	.141	.241	.310	2.01	2.29	2.90
8	.168	.272	.342	1.94	2.19	2.74
9	.193	.300	.369	1.88	2.11	2.62
10	.216	.325	.394	1.83	2.05	2.52
12	.256	.367	.435	1.75	1.94	2.36
14	.291	.402	.469	1.69	1.87	2.24
16	.321	.432	.498	1.64	1.80	2.14
18	.348	.457	.522	1.60	1.75	2.06
20	.372	.480	.543	1.57	1.71	2.00
30	.460	.560	.616	1.46	1.57	1.79
40	.518	.611	.663	1.39	1.48	1.67
60	.592	.675	.720	1.32	1.39	1.53
120	.699	.763	.798	1.22	1.27	1.36
∞	1.000	1.000	1.000	1.00	1.00	1.00

Table VII Resume' Of Confidence Limit Formulas

Parameter	Conditions	Point Estimate	Confidence Interval
μ	σ known	$\bar{X}$	$\bar{X} + z_{\frac{1}{2}\alpha}\,\sigma/\sqrt{N} < \mu < \bar{X} + z_{1-\frac{1}{2}\alpha}\,\sigma/\sqrt{N}$
μ	σ unknown	$\bar{X}$	$\bar{X} + t_{\frac{1}{2}\alpha}\,\hat{\sigma}/\sqrt{N} < \mu < \bar{X} + t_{1-\frac{1}{2}\alpha}\,\hat{\sigma}/\sqrt{N}$
$\mu_1 - \mu_2$	$\sigma_1 = \sigma_2$ known	$\bar{X}_1 - \bar{X}_2$	$\bar{X}_1 - \bar{X}_2 + z_{\frac{1}{2}\alpha}\,\sigma\sqrt{\frac{1}{N_1}+\frac{1}{N_2}} < \mu_1 - \mu_2 < \bar{X}_1 - \bar{X}_2 + z_{1-\frac{1}{2}\alpha}\,\sigma\sqrt{\frac{1}{N_1}+\frac{1}{N_2}}$
$\mu_1 - \mu_2$	$\sigma_1 = \sigma_2$ unknown	$\bar{X}_1 - \bar{X}_2$	$\bar{X}_1 - \bar{X}_2 + t_{\frac{1}{2}\alpha}\,\hat{\sigma}\sqrt{\frac{1}{N_1}+\frac{1}{N_2}} < \mu_1 - \mu_2 < \bar{X}_1 - \bar{X}_2 + t_{1-\frac{1}{2}\alpha}\,\hat{\sigma}\sqrt{\frac{1}{N_1}+\frac{1}{N_2}}$
μ_d $\mu_1 - \mu_2$	paired comparisons	$\bar{X}_d$ $\bar{X}_1 - \bar{X}_2$	$\bar{X}_d + t_{\frac{1}{2}\alpha}\,\hat{\sigma}_d/\sqrt{N} < \mu_d < \bar{X}_d + t_{1-\frac{1}{2}\alpha}\,\hat{\sigma}_d/\sqrt{N}$
σ^2		$\hat{\sigma}^2$	$\dfrac{\hat{\sigma}^2}{P_{\frac{1}{2}\alpha}} > \sigma^2 > \dfrac{\hat{\sigma}^2}{P_{1-\frac{1}{2}\alpha}}$
σ		$\hat{\sigma}$	$\sqrt{\dfrac{\hat{\sigma}^2}{P_{\frac{1}{2}\alpha}}} > \sigma > \sqrt{\dfrac{\hat{\sigma}^2}{P_{1-\frac{1}{2}\alpha}}}$

A DISCUSSION ON ROTATABLE DESIGNS

J. S. Hunter
Princeton University

Introduction

Recently a special class of experimental designs called "rotatable" designs has been proposed as being in some sense "best" when it is the purpose of an experimenter to explore an unknown response function using the techniques first proposed by G.E.P. Box. This paper is an attempt to describe what a rotatable design purports to do.

Some Background Material

Let us first assume that the errors made in repeatedly observing the response μ_i at the ith point in a factor space, are independent and Normally Distributed with a mean equal to zero and a variance σ^2. Then

$$y_{iu} = \mu_i + \epsilon_{iu} \qquad u = (1, 2, \ldots, N) \qquad (1)$$

Further assume that as we move from point to point in the factor space the errors remain Normally Distributed with mean zero and variance σ^2, and that between any two distributions of errors the covariance is zero.

Statistics are often formed from linear combinations of such observations. By "linear combinations" we mean that the statistic is calculated using a formula of the type

$$a_1 y_1 + a_2 y_2 + \ldots + a_N y_N \qquad (2)$$

where $a_1, a_2, \ldots, a_N$ are constants, and the observations $y_1, y_2, \ldots y_N$ appear only to the first power. For example, the average $\bar{y}$ is such a statistic since

$$\bar{y} = \frac{\Sigma y}{N} = \frac{1}{N}y_1 + \frac{1}{N}y_2 + \ldots + \frac{1}{N}y_N \qquad (3)$$

The variance of a statistic such as that shown in general in equation (2) is given by the formula

$$a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + \ldots + a_N^2 \sigma_N^2 \qquad (4)$$

and since the variances are all equal, by the formula

$$(a_1^2 + a_2^2 + \ldots + a_N^2)\sigma^2 \qquad (5)$$

The variance of the average $\bar{y}$ is therefore

$$V(\bar{y}) = (\frac{1}{N^2} + \frac{1}{N^2} + \ldots + \frac{1}{N^2})\sigma^2 = N(\frac{1}{N^2})\sigma^2 = \frac{\sigma^2}{N} \qquad (6)$$

Of course in practice the true variance of the errors, σ^2, is not known and must be estimated by calculating the statistic s^2 where

$$s^2 = \sum_{i=1}^{N}(y_i - \bar{y})^2/(N - 1)$$

Having calculated a statistic, and having estimated its variance, the t distribution can now be used to establish a confidence interval,

by substituting in the expression:

Statistic $\pm$ t (estimated variance of the statistic)$^{\frac{1}{2}}$ (7)

where the value of t depends upon the probability level at which the confidence statement is to be made and upon the degrees of freedom entering the estimate of the variance. Thus a confidence interval for the mean μ is given by

$$\bar{y} \pm \sqrt{\frac{s^2}{N}}. \qquad (8)$$

We now adopt the attitude that all postulated values of μ that lie within this interval are not contradicted by the data and are hence "acceptable" and that all postulated values of μ lying outside the limits are "unacceptable" or "significant". By "unacceptable" or "significant" we mean the probability that the given observations could have come from a population having this postulated value of μ is so small as to compel us to declare the postulate unreasonable.

Simple Linear Regression, 1st Order

Let us now suppose that an experimenter is studying the relationship between some response variable η and some simple controlled variable x. Initially nothing may be known about the relationship, the variable η being simple an unknown function of x, i.e.

$$\eta = f(x) \qquad (9)$$

As a first step in exploring this association between η and x, it is frequently assumed that the response η is a simple straight line function of x, i.e. that the relationship between η and x can be adequately represented by the <u>first order</u> model

$$\eta = \beta_0 + \beta x \qquad (10)$$

where β_0 is the intercept of the line and β is its slope. Since the observed value of the response y will never exactly equal η, the mathematical model actually used is

$$y = \beta_0 + \beta x + \epsilon \qquad (11)$$

For example suppose the following observations (the y_i) are recorded at the indicated settings of the controlled variable x, and that it is the decision of the experimenter to fit the first order model $y = \beta_0 + \beta x + \epsilon$. We then have

Controlled Variable x	Response Variable y
2	8
3	7
4	5
5	3
6	1

The least squares estimates of the coefficients β_0 and β are obtained from the following familiar formulas

$$b(\text{the estimate of } \beta) = \frac{\Sigma xy - \frac{(\Sigma x)(\Sigma y)}{N}}{\Sigma x^2 - \frac{(\Sigma x)(\Sigma x)}{N}} = \frac{Sxy}{Sx^2} = \frac{78 - \frac{(20)(24)}{5}}{90 - \frac{(20)(20)}{5}} = -\frac{18}{10} = -1.8$$

$$b_0(\text{the estimate of } \beta_0) = \bar{y} - b\bar{x} = \frac{\Sigma y}{N} - b\left(\frac{\Sigma x}{N}\right) = \frac{24}{5} - (-1.8)\frac{(20)}{5} = 12.0$$

The fitted equation is therefore,

$$\hat{y} = 12.0 - 1.8$$

where $\hat{y}$ stands for the predicted value of y at any chosen setting of x. Since $b_0 = \bar{y} - b\bar{x}$ another convenient way of writing the fitted equation is

$$\hat{y} = \bar{y} + b(x - \bar{x}) \tag{12}$$

$$\text{i.e. } \hat{y} = 4.8 - 1.8(x - 4).$$

The least squares fitted equation guarantees that the sum of squares of the quantity (predicted responses - observed responses) is a minimum,

i.e. that $\displaystyle\sum_{i=1}^{N}(\hat{y}_i - y_i)^2$ is a minimum.

The predicted values of the response calculated from the fitted equation are:

Controlled Variable	Predicted Responses	Observed Response		
x	$\hat{y}$	y	$(\hat{y}-y)$	$(\hat{y}-y)^2$
2	8.4	8.0	0.4	0.16
3	6.6	7.0	-0.4	0.16
4	4.8	5.0	-0.2	0.04
5	3.0	3.0	0	0
6	1.2	1.0	0.2	0.04
			0.0	0.40

We observe now that $\Sigma(\hat{y}-y)$ equals zero and that $\Sigma(\hat{y}-y)^2 = 0.40$.

The Analysis of Variance based on the first order model is:

	Sum of Squares	Degrees of Freedom	Mean Square	F Ratios
Total Crude Sum of Squares $= \Sigma y^2 = 148.0$		5		
Sum of Squares absorbed by $b_0 = \frac{(\Sigma y)^2}{N} = 115.2$		1		
Sum of Squares absorbed by $b = bSxy = 32.4$		1	32.400	F=243**
Residual Sum of Squares (by subtraction) 0.4		3	$0.1333 = s^2$	

As a check on the calculations we note that the residual sum of squares in the Analysis of Variance Table is equal to $\Sigma(\hat{y}-y)^2 = 0.40$. We observe further that the estimate of σ^2, the variance of the individual observations, is given by dividing residual sum of squares by its degrees of freedom, i.e. $s^2 = 0.1333$.

It should be remembered however that the difference between the predicted values $\hat{y}$ and the observed values y will measure experimental error and contributions caused by a choice of the wrong mathematical model. In this example it is assumed that the differences $(\hat{y}-y)$ are due to error alone. Accepting this estimate of s^2 as a valid estimate of σ^2 an F test of the mean squares associated with the coefficient b, with one and three degrees of freedom, is found to be highly significant and we conclude that the fitted model is an adequate representation of the unknown relationship $y = f(x)$ over the range of x studied.

A plot of the data and of the fitted first order model $\hat{y} = 12.0 - 1.8x$ is given below.

Figure 1
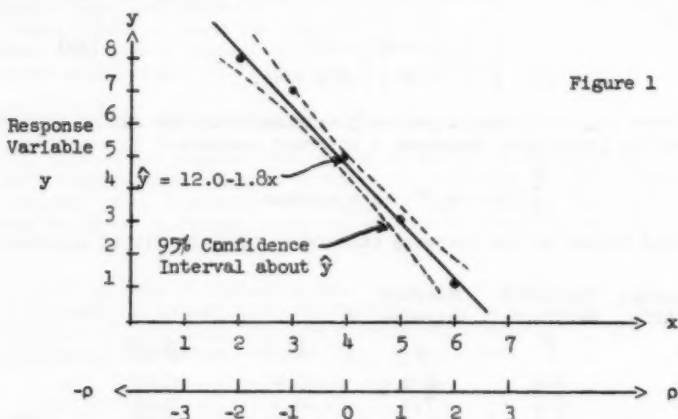
Now, as mentioned earlier, the confidence interval about any statistic formed from a linear combination of the observations, is given by the formula

$$\text{statistic} \pm t\sqrt{\text{variance of the statistic}}$$

The quantity $\hat{y}$ computed from the fitted equation for any value of x is such a statistic and we proceed now to construct a confidence interval about $\hat{y}$.

Since $\hat{y} = \bar{y} + b(x - \bar{x})$ the variance of $\hat{y}$ is

$$V(\hat{y}) = \text{Variance } (\bar{y}) + (x - \bar{x})^2 \text{ Variance } (b)^* \quad (13)$$

i.e.

$$V(\hat{y}) = V(\bar{y}) + \rho^2 V(b) \quad (14)$$

---

* We note here that the x's are controlled variables assumed fixed without error and have therefore no association with the errors made in observing the y's. Hence the quantity $(x - \bar{x})$ is simply a constant akin to the constants "a_i" given earlier in equation (2).

where ρ is the distance between the chosen value of x and the average $\bar{x}$. If we code the x's so that $\bar{x} = 0$ we may write

$$V(\hat{y}) = V(b_o) + \rho^2 V(b). \tag{15}$$

Thus a confidence interval about $\hat{y}$ for any setting of x is obtained using equation (7), that is, by calculating

$$\hat{y} \pm t \sqrt{V(\bar{y}) + \rho^2 V(b)} . \tag{16}$$

where $V(\bar{y}) = \dfrac{\sigma^2}{N}$ and we are given that $V(b) = \dfrac{\sigma^2}{Sx^2} = \dfrac{\sigma^2}{\Sigma(x_1 - \bar{x})^2}$. Of course in practice the variance σ^2 is estimated from the data.

In the example at $x = 2$, i.e. at $\rho = (x - \bar{x}) = -2$, the predicted response is $\hat{y} = 8.4$. The 95% confidence interval about the statistic is

$$8.4 \pm 3.18 \sqrt{\frac{0.1333}{5} + (-2)^2 \frac{0.1333}{10}}$$

or $8.4 \pm 3.18(.28)$ or 8.4 ± 0.9.

We can now obtain a confidence interval for $\hat{y}$ for all values of x. A plot of these intervals is also shown in Figure 1. We note from this figure that the variance of $\hat{y}$, and hence the confidence interval, is smallest when $\rho = (x - \bar{x}) = 0$ and symmetrically increases as ρ becomes larger.

Linear Regression, 2nd Order

Of course the experimenter may find that the first order approximation to the unknown response function is inadequate. One measure of the adequacy of a fitted model is to compare the quantity $\Sigma(\hat{y} - y)^2/(N - 2)$ against some measure of experimental error variance that is independent of the fitted model. Such measures are obtained by replicating observations at one or more settings of x. If the error variance calculated from replicated values is significantly smaller than that calculated by taking $\Sigma(\hat{y} - y)^2$ then the model is inadequate. In the event a first order model is found inadequate, or when the experimenter feels from the start that the response y gives a curve when plotted against x, a second order model can be fitted to the data, i.e. the model

$$y = \beta_o + \beta_1 x_1 + \beta_{11} x_1^2 + \epsilon \tag{17}$$

Some slight changes in subscript notation are to be noted. The controlled variable now appears as x_1 and its associated coefficient as β_1. The double subscript of β_{11} indicates that it is a coefficient associated with the second order term x_1^2. We cannot take the time here to show how to estimate the three coefficients β_o, β_1 and β_{11}. The calculations are a bit lengthy (but not difficult) and are fully described in several references (1), (2), (3). The fitted equation then is

$$\hat{y} = b_o + b_1 x_1 + b_{11} x_1^2 \tag{18}$$

The variance of $\hat{y}$ is

$$V(\hat{y}) = V(b_o) + \rho^2 V(b_1) + \rho^4 V(b_{11}) + 2\rho^2 \text{ cov.}(b_o, b_{11}) \tag{19}$$

where ρ is once again the distance from $\bar{x}$ to x. This formula for $V(\hat{y})$ requires a few comments. First, the values for x entering the

calculation must be coded so that $\Sigma x_i = 0$, i.e. so that $\bar{x} = 0$.
Secondly we observe that there is a covariance term occurring between
b_0 and b_{11}. This term will always be present since the estimated
coefficients b_0 and b_{11} as they appear in this model can never be
completely uncorrelated. References (1), (2), (3) tell how to compute
these variances and covariances. Knowing $V(\hat{y})$ we can now proceed to
establish a confidence interval about $\hat{y}$ calculated at different values
of x. Once again the size of $V(\hat{y})$ and consequently the size of the
confidence interval is smallest when $\rho = 0$, i.e. at the point $\bar{x}$ and
becomes progressively and symmetrically larger as the distance $\rho =
(x - \bar{x})$, plus or minus, becomes larger. The important point to note
about equation (19) is that variance of a prediction made at any point
x is solely a function of ρ, the distance from the center of the x
coordinate system.

First Order Rotatable Designs

Suppose now that a response y is an unknown function of two con-
trolled variables x_1 and x_2, that is

$$y = f(x_1, x_2) \qquad (20)$$

For example, the response may be the yield of a chemical batch process
and the two controlled variables the time and temperature of the reaction
As in the one dimensional case, the first step in exploring this function
might be to fit the first order model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \epsilon \qquad (21)$$

to the observations recorded at the different settings of the controlled
variables x_1 and x_2.

The most popular experimental design for obtaining estimates of the
coefficients in this first order model is the two level factorial design,
i.e. the 2^k factorial design with $k = 2$. This experimental design
requires that each controlled variable be fixed throughout the experi-
mental program at a high level and at a low level, and that all possible
combinations of high levels and low levels of the variables be run. For
example, in studying the relationship between the percent yield of a
process as a function of time and temperature the two level factorial
design would take the form:

Controlled Variable Levels		Experimental Design Levels		Response % yield
Time	Temp.	x_1	x_2	y
1 hour	240°	-1	-1	$y_1 = 1$
5	240°	1	-1	$y_2 = 6$
1	280°	-1	1	$y_3 = 7$
5	280°	1	1	$y_4 = 18$

The mathematical model is fitted to the experimental design levels
rather than the actual levels of time and temperature. The coding
mechanism associating the design variables x_1 and x_2 with time and
temperature are

$$x_1 = \frac{\text{time in hours} - 3}{2}, \quad x_2 = \frac{\text{temp } {}^{\circ}C - 260}{20}$$

The four experimental design points when plotted in the x_1, x_2

coordinate system form the vertices of a square as illustrated in Figure 2.
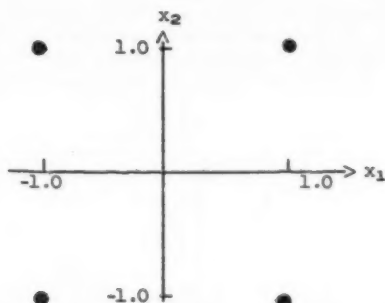
Figure 2.   Coordinates of the 2^2 Factorial Design

Estimates of the coefficients in the mathematical model are given by

$$b_0 = \frac{y_1 + y_2 + y_3 + y_4}{4} \; ; \qquad \text{thus } b_0 = \frac{32}{4} = 8.0$$

$$b_1 = \frac{-y_1 + y_2 - y_3 + y_4}{4}; \qquad \text{thus } b_1 = \frac{16}{4} = 4.0$$

$$b_2 = \frac{-y_1 - y_2 + y_3 + y_4}{4}; \qquad \text{thus } b_2 = \frac{18}{4} = 4.5$$

The fitted model is

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 \qquad (22)$$

$$\text{i.e. } \hat{y} = 8.0 + 4.0 x_1 + 4.5 x_2 \qquad (23)$$

Now the confidence in any value of $\hat{y}$ predicted at some setting of x_1 and x_2 depends, as shown, on the variance of $\hat{y}$. The variance of $\hat{y}$ at any setting x_1, x_2 is

$$V(\hat{y}) = V(b_0) + x_1^2 V(b_1) + x_2^2 V(b_2) \qquad (24)$$

The reader should be able to show quickly, using equation (5) that $V(b_0) = \sigma^2/4$; $V(b_1) = \sigma^2/4$ and $V(b_2) = \sigma^2/4$. Thus

$$V(\hat{y}) = \frac{\sigma^2}{4}[1 + x_1^2 + x_2^2] \qquad (25)$$

Using the Pythagorean theorem the distance out to any point x_1, x_2 is given by $\rho = \sqrt{x_1^2 + x_2^2}$ so that equation (25) can be rewritten as

$$V(\hat{y}) = \frac{\sigma^2}{4}[1 + \rho^2] \qquad (26)$$

Thus for the 2^2 factorial designs the variance of the predicted value $\hat{y}$ at the point x_1, x_2 is a function of ρ^2, the square of the distance to the point. Thus all predictions made at a constant distance from the center of the design have the same variance.

537

The condition that $V(\hat{y})$ be a function of the distance ρ is a necessary but not a sufficient condition to show that the 2^k factorial and fractional factorial designs are rotatable first order designs. For a design to be rotatable one other condition must be satisfied, namely that the variances of the coefficients must remain unchanged as the design is rotated. For example, suppose we rotate the 2^2 factorial design given in Figure 2 through an angle φ as illustrated in Figure 3.
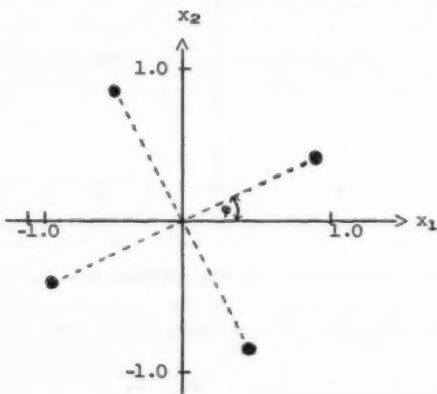
Figure 3.   A 2^2 Factorial Design Rotated through an Angle φ

What are the variances and covariances of the estimated coefficients when the observations are taken at these new points? It can be shown (ref. 4) that these variances remain unchanged, that is, $V(b_0) = \sigma^2/4$, $V(b_1) = \sigma^2/4$ and $V(b_2) = \sigma^2/4$ regardless of the orientation of the design. In addition to the fact that the variances remain constant we have seen that simultaneously $V(\hat{y})$ depends on ρ. Designs which meet these conditions are called rotatable designs. The 2^k factorial and fractional factorial designs, when used to estimate the coefficients in the first order model

$$y = \beta_0 + \sum_{i=1}^{k}\beta_i x_i + \epsilon \qquad (27)$$

are rotatable first order designs.   A very large class of rotatable first order designs exist, with the 2^k fractional factorial designs being, by far, the most widely used.

Second Order Designs

Very often the attempt to represent the unknown response function $y = f(x_1, x_2)$ by a first order model is unsuccessful and the experimenter then often attempts to fit the second order model

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_{11}x_1^2 + \beta_{22}x_2^2 + \beta_{12}x_1 x_2 + \epsilon \qquad (28)$$

The many uses of this second order model are mentioned in the references (1), (5), (6).

Let us first discuss an experimental design which is not rotatable

538

but which will provide estimates of all the coefficients in this model. The design in mind is the three level factorial design, i.e. the 3^k factorial where $k = 2$. The design requires that each of the controlled variables be held at a high (+1), middle (0) and low (-1) level, and that all combinations of levels and variables be run. The coordinates of a two dimensional, three factor factorial design and a diagram of the design are given below

x_1	x_2
1	1
0	1
-1	1
1	0
0	0
-1	0
1	-1
0	-1
-1	-1
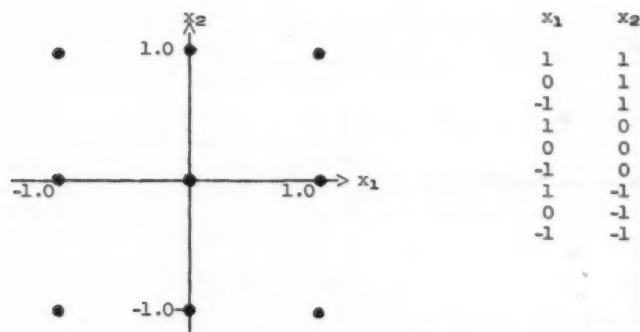
Figure 4.   Experimental Design Levels of a 3^2 Factorial Design

We note that the origin and scale of the design have been chosen so that $\Sigma x_i = 0$. If the response y is now recorded at each of these nine points there will exist sufficient numerical evidence to estimate all the coefficients in the second order mathematical model equation (28) to give the fitted equation

$$\hat{y} = b_0 + b_1 x_1 + b_2 x_2 + b_{11} x_1^2 + b_{22} x_2^2 + b_{12} x_1 x_2 \qquad (29)$$

Now it can be shown that

$$V(b_0) = \frac{5\sigma^2}{9} \qquad\qquad V(b_{11}) = V(b_{22}) = \frac{\sigma^2}{2};$$

$$V(b_1) = V(b_2) = \frac{\sigma^2}{6}; \qquad V(b_{12}) = \frac{\sigma^2}{4}; \qquad (30)$$

$$\text{Covariance}(b_0, b_{11}) = \text{Covariance}(b_0, b_{22}) = -\frac{\sigma^2}{3}.$$

The variance of $\hat{y}$ for the second order model is

$$V(\hat{y}) = V(b_0) + V(b_1)x_1^2 + V(b_2)x_2^2 + V(b_{11})x_1^4 + V(b_{22})x_2^4$$

$$+ V(b_{12})x_1^2 x_2^2 + 2\,\text{cov}(b_0, b_{11})x_1^2 + 2\,\text{cov}(b_0, b_{22})x_2^2 \qquad (31)$$

Substituting in the values of the variances and covariances of the coefficients and collecting terms gives

$$V(\hat{y}) = \sigma^2 \left\{ \frac{5}{9} - \frac{1}{2}(x_1^2 + x_2^2) + \frac{1}{4}(2x_1^4 + 2x_2^4 + x_1^2 x_2^2) \right\} \qquad (32)$$

A quick look at this equation for $V(\hat{y})$ should convince the reader that the variance of $\hat{y}$ at any point x_1, x_2 is not simply a function of the distance from the center of the design ρ but very much depends

upon the specific setting of x_1 and x_2. When this occurs we can quickly declare the design is not rotatable.

It is also interesting to observe what happens to the variances and covariances of the coefficients as this three level factorial design is rotated through an angle φ. For example, if $\varphi = 45°$ then the design given in Figure 4 becomes that shown below in Figure 5.

$$
\begin{bmatrix}
 & x_1 & x_2 \\
 & 0 & 1.414 \\
 & -0.707 & 0.707 \\
 & -1.414 & 0 \\
 & 0.707 & 0.707 \\
 & 0 & 0 \\
 & -0.707 & -0.707 \\
 & 1.414 & 0 \\
 & 0 & -1.414 \\
 & 0.707 & -0.707
\end{bmatrix}
$$
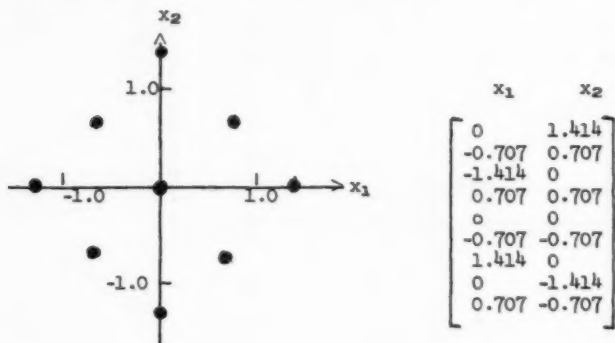
Figure 5.   Experimental Design Levels for a 3^2 Factorial Design Rotated through an Angle of $45°$

In this particular orientation the variances and covariances of the coefficients become:

$$V(b_0) = \frac{5\sigma^2}{9}; \qquad V(b_{11}) = V(b_{22}) = \frac{5\sigma^2}{16};$$

$$V(b_1) = V(b_2) = \frac{\sigma^2}{6}; \qquad V(b_{12}) = \sigma^2;$$

Covariance (b_0, b_{11}) = Covariance $(b_0, b_{22}) = \frac{-\sigma^2}{3}$;

Covariance $(b_{11}, b_{22}) = \frac{3\sigma^2}{16};$ \hfill (33)

Clearly the variances and covariances of several of the coefficients have changed and an additional covariance term, which was zero previously, has now appeared

Rotatability

We see therefore that if the $V(\hat{y})$ at a point x_1, x_2 changes as we move from point to point even when the points are all the same distance from the center of the design, the design is not rotatable. Viewed another way, if the variances and covariances of the coefficients in the model change depending upon the orientation of the design, the design is not rotatable.

In attempting to explore an unknown response surface using either a first or second order model, the experimenter will not of course know at which point in the experimental region his interest will ultimately lie. It would seem unfortunate then to use an experimental design which provided very small $V(\hat{y})$ at one point (and hence narrow confidence

540

intervals) and large $V(\hat{y})$ at some other point (and hence wide confidence intervals) even though both points were equidistant from the center of the system. Certainly such a design provides the experimenter with a rather astigmatic view of the response surface. However, if the experimenter uses a rotatable design the $V(\hat{y})$, and hence the confidence in any prediction, will be a function solely of how far the point of interest lies from the center of the design, the particular direction in which the point lies having absolutely no influence.

The 3^k factorials and fractional factorials do not provide second order rotatable designs, nor do the 2^k factorials. The 3^k factorials are however rotatable first order designs. This is quickly shown since if $\hat{y} = b_0 + b_1 x_1 + b_2 x_2$ then

$$V(\hat{y}) = \frac{\sigma^2}{9} + \frac{\sigma^2}{9} x_1 + \frac{\sigma^2}{9} x_2 = \frac{\sigma^2}{9}[1 + \rho^2]$$

and once again the variance of $\hat{y}$ at any point would be a function solely of the distance of the point from the origin of the design. A full description of the theory underlying the construction of a rotatable design is given in ref.(7).

A Rotatable Second Order Design

Let us now take a rotatable design such as the hexagon design with the center point replicated three times
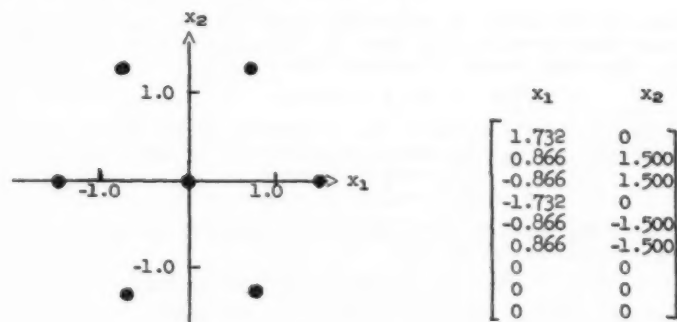
Figure 6.   The Hexagon Design

Note that the x's have been coded so that $\Sigma x_1 = 0$, $\Sigma x_1^2 = 9$, the total number of points. The fitted equation would, once again, be of the form given in equation (29). The variances and covariances of the coefficients are

$$V(b_0) = \frac{\sigma^2}{3}; \qquad\qquad V(b_{11}) = V(b_{22}) = \frac{5\sigma^2}{54};$$

$$V(b_1) = V(b_2) = \frac{\sigma^2}{9}; \quad V(b_{12}) = \frac{4\sigma^2}{27}; \qquad\qquad (34)$$

$$Cov(b_0,b_{11}) = cov(b_0,b_{22}) = \frac{-\sigma^2}{9}; \quad cov(b_{11},b_{22}) = \frac{\sigma^2}{54}$$

Thus, using the second order model we get for the variance of $\hat{y}$

541

$$V(\hat{y}) = \sigma^2 \{ \frac{1}{3} + \frac{1}{9}(x_1^2 + x_2^2) + \frac{5}{54}(x_1^4 + x_2^4 + 2x_1^2 x_2^2) \} \quad (35)$$

or, letting $\rho^2 = x_1^2 + x_2^2$

$$V(\hat{y}) = \sigma^2 \{ \frac{1}{3} + \frac{1}{9}\rho^2 + \frac{5}{54}\rho^4 \} \quad (36)$$

Here we see that the variance of any predicted value is a function solely of how distant the point is from the center of the design. Furthermore, it can be shown that the variances and covariances of the coefficients in the second order model do not change when the design is rotated through an arbitrary angle φ (ref.7). Hence, the hexagon design is a rotatable design.

Conclusion

Early in this paper it was shown that if we fitted the simple first order model $y = \beta_0 + \beta x + \epsilon$ (and took the precaution to code the x's so that $\Sigma x = 0$) then the variance of a prediction made at a distance ρ from the point $x = 0$ is

$$V(\hat{y}) = V(b_0) + \rho^2 V(b)$$

If a first order rotatable design is used to fit the first order model

$$y = \beta_0 + \sum_{i=1}^{k} \beta_i x_i + \epsilon$$

then regardless of the number of controlled variables $x_1, x_2, \ldots, x_k$, it is similarly true (provided that each x_i is coded so that $\Sigma x_i = 0$ and $\Sigma x_i^2 = N$, the total number of points) that

$$V(\hat{y}) = V(b_0) + \rho^2 V(b_1).$$

In this instance ρ is the distance from the center of the design to the point $(x_1, x_2, \ldots, x_k)$ at which the prediction is made, i.e.

$$\rho = \sqrt{x_1^2 + x_2^2 + \ldots + x_k^2}$$

Similarly it was shown earlier that if the second order model $y = \beta_0 + \beta_1 x_1 + \beta_{11} x_1^2$ is fitted then

$$V(\hat{y}) = V(b_0) + \rho^2 V(b_1) + \rho^4 V(b_{11}) + 2\rho^2 \text{cov}(b_0, b_{11}).$$

If a second order rotatable design is used to fit the model

$$y = \beta_0 + \Sigma \beta_i x_i + \sum_{i=j}^{k} \Sigma \beta_{ii} x_i^2 + \sum_{i<j}^{k} \Sigma \beta_{ij} x_j + \epsilon ,$$

then regardless of the number of variables $x_1, x_2, \ldots, x_k$, it is similarly true that

$$V(\hat{y}) = V(b_0) + \rho^2 V(b_1) + \rho^4 V(b_{11}) + 2\rho^2 \text{cov}(b_0, b_{11}).$$

It is this element of simplicity, i.e. the one to one correspondence in obtaining $V(\hat{y})$ using regression equations with one or with k controlled variables, that identifies as unique the rotatable designs.

Bibliography

(1) Davies, O.L, (Editor), <u>The Design and Analysis of Industrial Experiments</u>, Box, G. E. P., Connor, L. R., Cousins, W. R., Himsworth, F. R. and Sillito, G. P., Oliver and Boyd, London, 1954.

(2) Anderson, R.L. and Bancroft, T.A., 1952, <u>Statistical Theory in Research</u>, McGraw-Hill, New York.

(3) Hunter, J.S., Notes for the Course "The Box Method of Experimentation", Chemical Division, A.S.Q.C., 1956.

(4) Box, G. E. P., 1952, "Multifactor Designs of First Order", <u>Biometrika</u>, <u>39</u>, p. 49-57

(5) Box, G. E. P., and Wilson, K. P., 1951, "On the Experimental Attainment of Optimum Conditions", <u>J.R.S.S., Series B</u>, <u>13</u>, p.1-45.

(6) Box, G. E. P., 1954, "The Exploration and Exploitation of Response Surfaces", <u>Biometrics</u>, <u>10</u>, p. 16-61.

(7) Box, G. E. P. and Hunter, J. S., 1957, "Multifactor Experimental Designs", <u>Ann. Math. Stat.</u>, <u>28</u>, No.1, p.195-241.

PRODUCTION RELIABILITY

Richard A. Carlson
Napco Industries, Inc.

The question of Reliability has received much attention in engineering design groups, QC staff groups, component evaluation groups, but the department at the end of the line "Production" seldom is considered. Much effort has been expended in some areas to achieve Reliability in production on specific assemblies or a product. Little effort has gone into the general application of increasing Reliability in production. Due to the nature of the beast, concentration in specific narrow areas won't raise the overall levels of Reliability.

Production Reliability can be achieved if the desire to improve is implanted in all levels of production management. It is not just another technique like QC. It cannot be delegated to a staff assistant. By themselves production engineers, QC engineers or the janitor cannot insure the success of the program. But all levels from the production workers through the foreman to the plant manager must have the Reliability religion. The staff groups servicing production must also be infused with this same fervor. Then you can expect tangible results from reliability applied to production.

The answers to three questions will supply us with ideas necessary to apply reliability to production. Basic understanding by everyone as to Why, Where and How of production reliability is vital to the success of the program.

Why

For those engaged in strictly military work part of the why is obvious. More and more contracts are specifying a reliability probability. The application and translation of this probability into an equipment test program can be quite difficult. The increasing complexity of systems and equipment is another factor involving Dr. Robert Loesser's famous cry "p overall". High maintenance costs coupled with high failure rates bring the customers, either commercial or military, screaming back to the sales department demanding some changes. Equipment failure at critical times never makes a customer happy. The loss of a mission, lives or equipment brings a very serious picture into focus.

These comments are intended to show the need for increased reliability. This is applicable to all areas but intended to be used here as reasons for the need of increased production reliability. A need must be shown to people before they become interested enough in a problem to take action.

The notation "p overall" is a concept that is very important in the understanding of where and how production reliability can best be achieved. The probability p overall is merely the product of the probabilities of all of the individual components. This meaning is expressed in terms of the end product performance. An example of a p overall of 80% in a unit of 400 components allows only one failure in 1800 units of each of the components. Thus the understanding of this concept is necessary before proper steps can be taken to increase

production reliability. This concept has an effect on the method and direction taken in attacking problems, to the extent even of the sequence of the problems.

Where

The problem of where to look for production reliability begins with a definition. The normal definition of reliability is "worthy of confidence." This applied to production brings to the focus that, what is built must be worthy of confidence. This begins with the state of mind of all personnel involved. Everyone must be convinced that the products being built are to be the best of everyone's effort. The product is the best design, built with the best tools by the most highly skilled craftsman in the industry. This feeling must permeat all personnel. Thus the definition of production reliability is "a state of mind." The attitude of all personnel towards workmanship or craftsmanship, a pride in doing a fine job is a prerequisite of production reliability.

The discussion on workmanship leads us to other areas. It is necessary to understand the team concept of working together. The team consists of all personnel in engineering and production as well as staff groups assisting. The line worker must feel the same pride as the engineer. Each must feel that the other is a necessary part of the team, with the product as the common bond.

Before reliable production of a product can occur, there must be a reliable product. The design must be such that it results in a highly reliable product, in short the engineer must have done an excellent job. The problems of reliability barriers, component evaluation, component variabilities must have been solved with the best compromise. The reliability deficit between growth of equipment complexity and growth of component reliability is growing wider. A reliable design must come first. This design and its components must be tested to failure. The reliability boundaries and safety margins must be checked and evaluated again and again. Due to our friend p overall the usual concepts of specifications are not good enough. The interaction of component variability, variables in adjustment and alignment, interaction of assemblies grow more complex each day. The state of the Art is not yet mature in these areas.

Even as a well tested reliable design arrives in production new yardsticks must be used in processing and manufacturing. The usual techniques are not adequate. The ordinarily accepted QC concepts are not adequate. Processing cannot be compartmentalized. Effects of a later process can negate satisfactory early processes. Careful selection of components can be ruined by later immersion in hot solder or paint drying ovens. Care must be exercised in handling and in the attitude towards errors. Each person should feel that the best is none too good. If an operator finds an error missed by inspection or other workers it must be culled out and rejected.

Quality Audits by engineering of finished products in final shipping must be continuous. Changes made in design and processes must be evaluated in test programs. One aircraft equipment manufacturer discovered to his horror that 100% of his fuel gage equipment began failing. Original models had passed all airborne and environmental testing. But over an eighteen month period a cumulative effect of many minor changes had taken their toll on the equipment's ability to pass the

specifications. Each small change had been checked, but few tested. Now this manufacturer has a Tight Quality Audit System on shipped goods stock.

Analysis of Where on production reliability shows that the engineering and production team must first have a reliable design thoroughly tested and evaluated with continuous evaluation of production units.

How

The How of production reliability must again cover more than just the manufacturing areas.

The first and most important part of production reliability is education. As has been repeated before in this paper, everyone must be quality and reliability conscious. This feeling has to be coupled with training programs stressing workmanship. Quality and Reliability meetings are held periodically like Safety meetings. That is why all levels of management must be convinced the reliability can and must be achieved. Other training programs for workers covering elementary QC principles and work simplification keep them informed and interested in what's going on. The use of production handbooks similar to the Western Electric type assist in new worker training and in getting consistent performance of many assembly process hand techniques.

Design analysis for producibility by production and tool engineers is necessary early in the design phases. Ease of manufacture and cost reduction can be accomplished if there is good relations in the engineering team. Each must respect the other and realize that both production and design engineers are necessary. There statistical quality control techniques are useful when applied to models or short runs.

The processing must be carefully considered to avoid damage or undue stress to components. The processes must be tested and analyzed from many angles. Tooling and inspection test equipment needs detailed analysis. There again statistical quality control techniques are of great assistance in model and trial run analysis. The use of production environmental testing is controversial. Care must be exercised else more troubles will be added into a product than errors called out. Still the best test for a cold solder joint in my opinion is a nice cold bath in air in a freezer at -50°C, then bring to room temperature abruptly. Excessive testing may shorten the life of the product.

Close liaison must be maintained between component engineers, purchasing, production engineering, inspection, QC and production control to see that component problems with vendors do not get out of hand. Attitudes towards specifications, component testing at vendor and home plants must change. Vendor education is a must. Materials handling of components, both shippers and in plant, must be under careful control. In missle work 63% of failures have been due to component failure.

Subcontractor problems are similar to that of component vendors complicated by the fact of performance specifications of assemblies. It must be emphasized that a subcontractor is an extension of your own plant. You cannot accept conditions in his plant that you wouldn't accept in your own. Again education is necessary. In this respect production engineering must have firm control of subcontractors and purchasing and production control.

547

The manufacture of piece parts is fairly straight forward, but it too can have pitfalls. Fortunately experience to date has shown that the reliability of structures is good. There have been no missle failures due to air frame failure for example.

The Assembly areas offer fertile fields for problems to occur. Here supervision and training are the keys. Union relations must be considered and can very often be extremely helpful if approached properly. The functions of inspection and verification must be carefully understood and delineated. Inspection is performed at definite points with written procedures and accepted equipment. Verification is performed by operators on their own or previous work, where errors are discovered as operations are being performed. The unit is rejected in either case as errors are discovered. Both types of rejects are analyzed by QC. The emphasis is returned again on education and training of supervision and working force. A foreman told me once that he was happy with one bad solder joint in 5,000 joints. I pointed out that this was a p overall of about 60% and wasn't acceptable. A rate of one bad in 100,000 or better should be his goal. He left with his head down like a bulldog. Some surprising things can turn up in production evaluations when looking at it from new aspects.

In the QC and Inspection area the usual techniques of quality control are not completely adequate. Failure analysis by engineers of production rejects, test failures and field failures is also vitally necessary. An example of failure analysis is that of printed circuit chassis. By using punched card systems of lot identity of materials, components and assemblies it is possible to examine the history of any one printed circuit assembly. Does it have an excessive number of rejects or replacements, do the lots from which the components come have poor reject rates, were there process troubles during the assembly of the unit? Thus it is possible to look at a finished assembly that has passed inspection yet say that due to various conditions of its components or processes that it might have a high probability of failure, therefore reject it. The Quality Audit as described before is another link in the chain of reliability.

The lot identity program as described above as handled by the material analysts is a bit complicated. The advantages in the initial production runs far outweigh disadvantages. The knowledge gained is extensive. These must be handled with common sense. They can include raw materials, components and processes. A very complete picture can result.

The last item for discussion is costs. The initial costs are high and they are difficult to justify now on equipment which has no maintenance or failure history. But the attempt must be made by management to budget reliability costs into a program. They then can be handled like any other item of cost. It has been discovered that in production other economies result from reliability work. Normal results are low reject and rework rates in production. Component sorting by semi automatic means has also resulted in lower reject and rework rates. Dollars can very often be hooked to this sort of savings. A final remark can be made concerning costs, you can compare the cost of reliability with the high cost of failure.

In summation it can be said that for military requirements and commercial customers too that you must become reliability conscious.

548

You might as well get on the band wagon but with some common sense.
Reliability isn't too rough to get into if you aren't too hidebound to
begin with. By using ingenuity, common sense and salesmanship highest
reliability can be achieved in production - it can be maintained - and
it can be done at a reasonable cost.

ROLE OF RESEARCH IN QUALITY IMPROVEMENT

Richard D. Wells
Assistant Director
Fabric Research Laboratories, Inc.

My comments to this group today will be made with the assumption
that all of us here accept both Research and Quality Control as vital
factors of business success. However, such an assumption would not be
as certain with all textile audiences. There is still a strong senti-
ment, in this primarily production- and price-oriented industry, that
product "qualities" have only secondary significance. While there are
notable exceptions to such a sweeping generality, it is a fair asser-
tion that in the textile industry at large, research and quality control
are still classed as "supplemental services", and have not assumed full
stature as contributors to corporate policy and strategy.

This statement is not made from a sense of frustration, nor is it
made in a derogatory sense. My point is that perhaps neither executive
management, nor ourselves, have considered research and quality func-
tions in the full potential frame of reference. It is my purpose to
suggest some of the opportunities to extend the roles of research, and
of quality control, to good advantage. These opportunities lie, for
the most part, in the closer collaboration of the two functions.

In contrasting research with the usual limited role of quality
control, one can say that in the first is the science of what might
be, while the latter is the art of tolerable operations with what now
is. Research is optimistic in the sense that it always assumes better
things are possible. Quality control is largely premised on the belief
that perfection is improbable. Both attitudes have their place, but
must be kept in balance. A philosophical comment might be that close
collaboration would have value, if for nothing else, in tempering re-
search with a bit of practicality, and in prodding quality control to
a bit more dissatisfaction with the status quo.

It has been helpful to me in organizing thoughts on this subject
to refer to basic definitions. We are, I am afraid, rather prone to
sloppy semantics, being so accustomed to narrow or specialized meanings
that we tend to overlook the implication of the terms we use. As has
been pointed out by more able scholars than I, human thinking processes
are bound entirely by words and meanings. Corporate thinking is nowise
different.

We see so-called "Research Departments" which, whatever their value
may be in other services, make no pretense of research other than in
name. There are capable Quality Control Departments, perhaps fulfilling
a real need, but which may have very little to do with "quality", and
have no semblance of "control". The error is to assume, since the terms
appear on the organization chart, that the research and quality control
needs of the company are covered. This may be true in part, or not at
all. Very seldom are these functions thought through and defined in
specific terms in the same sense as are the other business functions
such as production, sales, engineering, or accounting. At this time
when so much soul searching is going on to seek causes for the ills of
the textile industry, it may not be out of order to suggest that inade-
quate understanding of words may be a contributing factor.

I have confidence, at least that proper terminology will clarify issues and aid the diagnosis.

Following Webster, we shall consider research as "careful search or inquiry, usually critical and exhaustive investigation or experimentation, having for its aim the revision of accepted conclusions in the light of new discovered facts". The "conclusion" of the research process is not necessarily an abstract theory, nor, in the other extreme, a new product idea. It may be a very down-to-earth conclusion, as for instance that a diligently sought after improvement in a product or process may be inherently impossible, or actually meaningless, or perhaps even undesirable. It might be that in a certain quality control problem the variance at A is positively related and calculable from observed variance at B, and hence requires no separate testing. It might be, from an operations research study, that on a long term basis the quality loss penalty of a proposed new system, work assignment, or cleaning and maintenance schedule, would more than offset all positive advantages. No matter under what title a person may draw his pay, he is conducting research if he is deliberately seeking facts on which to base logical revised conclusions. The "revision" may be something new, something changed, or the confirmation or denial of a hunch or rule-of-thumb.

Examining the definition of "quality", we find three distinct meanings pertinent to our subject. Quality can refer to any characteristic, good or bad, intrinsic or incidental, as a describable and perhaps measurable property. We can speak, for instance, of typical "textile qualities", as distinct from paper or plastic film qualities. These are the characteristics that make something what it is, that make it desirable for some functions, and perhaps undesirable for others. In a sense, this meaning can be taken as the "qualitative" sense of quality.

The second meaning, one might say the "quantitative" sense of quality, relates to grade or relative degree. It represents a ranking on some scale of values. We speak of high quality, low quality, or below standard quality, always assuming, of course, that the one to whom we are speaking knows and accepts our frame of reference.

The third meaning of "quality" connotes general excellence. In this sense "quality" is a rather snobbish term, and those who use it generally avoid specific reference to individual attributes supposedly contributing to that excellence. Perhaps this meaning can be better understood as the antithesis of shoddy, skimpy, uneven, or unreliable. But even that is not valid in all cases. A "quality" tweed to many minds is the crudest, harshest, and most amateurish looking one that can be found. In women's hose, it isn't a sign of "quality" to last more than a few days of wearing. But whatever the basis may be, technical or entirely subjective, quality excellence is primarily an impression in someone's mind gained at least in part from the assumed attributes of the item. This sort of observation may seem rather remote from statistical quality control, or technical quality improvement, but it is worth a pause to consider that a narrow CV chart never sold a yard of goods except as somebody had been conditioned to consider that a sign of quality excellence.

The word "control" is another term that has in technical usage taken on an unduly restricted meaning. There are actually two quite

distinct implications. We commonly think of control only in the regulative sense of checking, adjusting, and classifying. A main purpose of these remarks is to focus attention on the other implication of "control", namely, the directive or guiding role. As useful and necessary as the regulative function may be, perhaps the biggest opportunity for benefits from quality control is in the directive role. In essence, that means decision making responsibility for the pattern to be followed rather than merely the policing of the operations to assure conformance to the established pattern. The term "Quality Improvement" as used in the title of this paper refers to one phase of this directive control. While on the subject of definitions, however, it should be noted that "improvement" has a meaning, "to turn to good account", (i.e., improving one's time) which is fully as appropriate to our general topic as its common connotation of enhancement, or increase. As will be discussed later, improvement may at times be in the lowering of a standard, if, considering all other conflicting factors, such reduced quality is more appropriate for the need.

Taking the several meanings of "quality" and "control", in their possible combinations, we find a much more comprehensive view of quality control than that usually envisioned. Though I would not maintain that operationally these are six wholly separate functions, the following table is a useful reminder of objectives and possibilities:

		Control	
		Directive	Regulative
	Character	Design	System
Quality	Grade	Standards	Operations
	Excellence	Policy	Promotion

The terms inserted in the boxes are suggestive rather than definitive, and the balance of my remarks will be directed to the exploration of these six terms.

Design - Directive Control of Quality Character

Design involves the determination of what properties the product is supposed to have to make it useful and desirable. Quality design is just as important to raw materials and intermediate products as it is to the final items of commerce and use. For instance, the characteristics of a picker lap have no significance whatever except as they render it suitable or unsuitable as a feed for a card which in turn delivers a product with certain requirements determined by the next operation. This may seem very obvious, and yet I think there is a growing tendency in quality control work to look backward instead of forward, to measure and judge qualities with reference to the capabilities and variations of the process from which they are derived, rather than by relation to characteristics of significance in the end use. In the case of the picker lap just mentioned, we customarily note variations in weight which are to some extent controllable, and devise a system of checks and adjustments to keep the average weights to a prescribed level and the variations in a workable range. Yet, a deeper investigation might show that the state

of fiber orientation and entanglement, or of fiber strain and rupture, also materially affect the suitability of the lap on a feed for the card. Unless one knows for a fact that these other "qualities" are either constant or insignificant, it is rather naive to assume that if the weights are satisfactory all is well. Some of these other qualities may be obscure and difficult to characterize except by refined research techniques. But it would be rather surprising if, in the last analysis, it was found all that counted in these textile intermediates was weight.

As another example, consider the quality design of a fabric for a certain industrial application where high strength is a requirement. That would seem to be a straightforward proposition. Yet what do we really mean by "high strength", in any particular case? Is it average strength in one-time loading, is there a critical spacing of strength-giving components, is it retention of strength in repeated loading or flexing, is it a unidirectional or a multidirectional requirement, or is it only ability to resist distortion under loads much less than ultimate strength? It has not been customary to consider that such factors are in the province of quality control. The usual procedure is to take some sample that has been "accepted" and control the production system to duplicate it within reasonable tolerance according to obvious and conveniently measured characteristics. The characteristics being duplicated, however, may be only incidental and quite insignificant in terms of true requirements. It is not at all uncommon, particularly in industrial application, to find two items which conform equally in normal specifications, but one proved superior to the other in service. There is no mysterious anomaly in this, but only failure to properly determine quality character and direct the design specifically to have that character. It might very well be, as perhaps in a fabric for reinforcing diaphragms, that a rather low average strength could be tolerated as long as there were no localized areas of weakness. It should certainly seem important for the quality controller to know that requirement so that the production and control system be set to eliminate the low extremes even if the averages or the statistical variability has to be sacrificed to some extent in achieving that objective.

I can think of an instance of a company pioneering in a new type of fabric for an important industrial use. This involved novel manufacturing techniques which were difficult to master. Early production was accepted only under the premise that subsequent experience would correct a certain apparent "fault" still under criticism. As it happened, a potential competitor got wind of the situation, devised a system which eliminated the fault, and seemed set to scoop the field while the original producer was still struggling. It was a technical triumph, except for the fact that his improved material didn't work at all in this application. Later analysis showed that the "fault" was the key characteristic for success in this critical use.

Character design is also important even with standard staples of presumably non-critical materials. As a simple and yet dramatic case, there was a mill of diminishing fortunes suddenly inspired to out-modernize its competition and win its way back with a premium grade of goods. This objective required a more uniform yarn than was customary to this trade, and the new equipment and systems were chosen, after quite extensive study and evaluation, with that as the primary criterion. Initial samples were beautiful. However, the desired characteristics had not been fully and properly defined. This was a fabric of contrasting

554

warp and filling color. The new yarn was so perfect and uniform within
the bobbins that the very slight bobbin-to-bobbin variations caused
shade bands of surprising contrast. The distribution of this irredu-
cible variation within the total system was such, since sorting and
channeling were unfeasible, that the mill was turning out many cuts of
goods of truly exceptional quality, and almost as many which it would
be only charity to classify as commercial seconds. There was nothing
in between. A certain amount of eye-distracting short-term variability
would have been far preferable to the perfection that had been achieved
at great expense and effort. The same, or even an appreciably greater
degree of bobbin to bobbin variation would have been indistinguishable
if the individual bobbins had been less uniform and solid in appearance.
A critical determination of what was important in this type of goods,
in other words, a definition of desired quality character, would have
prevented this fiasco. The advance evaluation would have then been by
reference to the specific requirements rather than to what the ordinary
mill or regulative quality control man would consider general criteria
for good uniform yarn.

It is obvious, I think, that this Design phase of Quality Control
calls for research attitude and research methodology. Facts must be dug
out by analysis and critical experimentation as basis for sound conclu-
sions as to what qualities should be controlled. Assumptions or over-
sight may be dangerous.

System - Regulation Control of Quality Character

Once the desired characteristics are determined it is essential to
devise a system by which they may be checked. That is so logical a step
that it would seem to require little mention. Unfortunately, it seems
so patently simple that it sometimes gets no attention at all. The two
rules for proper regulation of quality character are actually versions
of the same principle. The one is not to confuse the index with the
property, and the other, not to assume an intrinsic correlation of two
properties unless such is proved certain.

Take, for instance, the matter of uniformity of yarn as a charac-
teristic. We have mentioned previously a case where the wrong index was
used insofar as short, medium, and batch uniformities were concerned.
But take another look at the same problem and we see further that in this
case, as in most other textile cases, it is bulk uniformity and not weight
uniformity that is the real characteristic to be achieved and controlled.
Yet most of our uniformity measuring devices are based on weight, or a
direct index of weight such as capacitance. Even those devices which do
sense bulk do so under pressure and confinement designed to minimize
density of bulk effects, and thus presumably indicate weight. The cor-
relation between weight and bulk may be good enough for some purposes
and not for others. I recall the withering scorn addressed to a quality
control spokesman defending his product in an argument, which must be
paraphrased for public utterance as, "Young man, I am not buying your
control charts. My eyes tell me the yarn is lumpy, and you can keep it".
We have, on the other extreme, the hard-shell who has no faith in any
instrument determination and trusts only his eye. Some of these gentle-
men have in fact very discerning judgment and on the seriplane boards
can consistently detect differences too subtle for present instruments.
Yet they can be equally wrong, for the final criterion of uniformity is
that of the effective bulk under the constraints of the woven or knit

structure, which are quite different from the conditions on the viewing boards. Some types of yarn variability are almost canceled out as far as visible effects are concerned under the constraints of the final fabric, while other types of variability where stiffness is a correlative factor are augmented in apparent effect due to the influence on crimp balance. This subject of yarn uniformity is highly complex as you well know, and I do not intend to launch into a thesis but only to raise the question as to whether some of our ingenious and statistically sound quality control systems are based on the factors we should be measuring and controlling.

Another case, perhaps more remote from your field but worth mentioning as an extreme case, is that of a cotton mill which ran into great difficulty with cotton from a new year's crop, in that it gummed up and lapped the draw rolls. It was a sticky problem, and no joke, for downtime was mounting to disastrous proportions. The troublesome substance seemed sugary in nature, as perhaps from incomplete synthesis of cellulose due to peculiar weather in that crop year. With that clue a system was set up to check each incoming bale by a reducing-power test for sugars. The "sweet" bales were flagged and proportioned into the laydown mixes in low percentages, and the problem was brought quickly under control. The daily sugar tests were checked as carefully as if the mill were a diabetic clinic. This quality control operation was followed through the next crop season and the next with great apparent success, for though sweet bales were still showing up, no trouble was encountered following this system. However, for some reason the laboratory once fell behind on the tests and for several days mixes were made up without regard to sweetness. When the reports finally came through it was realized that several of the mixes had been "loaded" with sweet bales. The mill didn't know the difference. There was no sticking. Here was a case of double confusion, for a reducing test is at best only a general index of sugar content, and sugars were only in part an index of stickiness. What had proved a highly valuable measure under a certain set of conditions was inappropriate and misleading under other circumstances. Correlation had been assumed, but not proven, and the quality control system adopted without adequate research to establish valid premises for the test procedures to insure their pertinence to the desired characteristics.

Standards - Directive Control of Quality Grade

This function begins to get a little closer to what is ordinarily considered "quality control". But there are two research steps in this that are quite frequently missed.

In discussing the matter of Design, we were in a sense working toward what might be called the "ideal specification" of qualities we would like to attain. The first phase in establishing standards is to set this ideal specification out in balance and in quantitative terms. In most frequent cases there will be found conflicts based on fundamental factors. For instance, a requirement for thickness might have to be reconciled with that for low weight, porosity, strength, or drape. Based on relative values of the several requirements, an ideal target is set. Upper and lower values should be established, where possible, indicating limits beyond which that characteristic is either unacceptable on the one hand, or is in a meaningless range on the other. One might, for instance, determine that below a minimum of 20 pounds strength

the items would fail to function at all, and above 30 pounds no benefit
would accrue from the excess. In the aforementioned case of the con-
trasting color fabric, it could have perhaps been determined that a CV
for a certain gage length of 15% would disqualify the merchandise, but
any improvement below 5% would be undetectable in that particular fabric.

The next step is the determination of the potential specification
based on the inherent capabilities of the materials and processes with
which we must work. It is a "maximum" standard in the sense that it is
the best we can hope to do if everything goes as perfectly as possible.
Failure to research these limits has caused untold grief in textiles
and I imagine in other industries as well. Unless the potential maximum
specification falls safely and comfortably in the ranges of the ideal
specification, the mill would be well advised to turn its attention to
other affairs. All too often a commitment is made without regard to the
potential specifications, in some vague hope that the production and
quality control crowd can somehow achieve the ideal specification.

An excellent example of this point is in the blending of two fibers.
Suppose a 50/50 blend is the desired norm. An experimental checkup might
show for the ideal specification, that variability (expressed as the CV
of the ratio of the percentage of the two fibers in a given gage length)
in excess of 10% would be unacceptable in view of shade, luster, napping,
or other requirements. It might also show that below 5% there were no
discernible differences.

The potential specification might show that in the production sys-
tem to be followed there were inherent random elements which precluded
the chance of operating within the ideal specifications no matter what
control was exercised, or that the potential was so close to the ideal
specification tolerance that the additional variation of non-inherent
factors would throw much of the product beyond the acceptable limit.
For instance, in a rather intensive statistical research study on two-
component blending made a year or two ago it was shown that for the
chosen fibers in their lengths and deniers a CV for blend ratio of below
8% for a certain gage length was a theoretical improbability. As it
happened in this case this limit held for both hopper and draw blending.
If the ideal specification called for a 10% maximum, it meant that
things were too close for comfort.

These two theoretical specifications, the ideal and the potential,
may at first seem only abstract concepts of little value. I would cer-
tainly grant that since this is the textile business we are talking
about, and not guided missiles, there is seldom time or funds available
to do all the research necessary to establish complete and precise speci-
fications of this sort. However, the key characteristic generally can
be singled out and determined to a useful approximation. The value lies
in putting the final operating quality control system in proper per-
spective. If, for instance, it is known that even the best one can do
in achieving a certain characteristic is close to the ideal tolerance
limit, the working tolerances should be adjusted accordingly and cor-
rective attention concentrated on factors influencing that character-
istic. If, on the other hand, the quality control staff knows only the
intended norm, the usual tendency is to set tolerances solely on statis-
tical considerations of the product variables actually observed rather
than on variance from limits inherent to the system. The uses of the
potential specification or standard are in two directions. On the one

557

hand, it avoids plans and commitments which are impossible of attainment. On the other, it gives us a much more realistic picture of the efficiency of our controls, as in the blending problem for instance. By ordinary comparisons an observed CV of 10% might be considered only slightly less uniform than a CV of 9%. With reference to a potential standard of 8%, however, it is recognized that the 10% is twice as far off the beam, or out of control, as the 9%. The main purpose of the potential standard is to distinguish the controllable from the uncontrollable, and furnish a meaningful base against which to measure product conformance and in turn, the efficacy of our controls.

The working specification is the one with which we are most familiar. It is based largely on previous or current experience. The danger is that if the ideal and the potential specification are not kept in mind the working specification becomes something sacred and immutable. The conditions which dictated the compromises of the working specification, at the time it was drawn up, may change, perhaps by an incidental change somewhere along the line, perhaps by deliberate corrective action. This may be from a basic modification of the system which in effect alters the potential specification and gives opportunity to improve the working specification even without any better "control". It may be by a new development allowing closer control of the controllable variables, with no change in the ideal specification itself. The guiding rule is to consider every working standard as tentative and subject to revision as research points the way to improvement.

Operation - Regulative Control of Quality Grade

This is the field you know best and I shall not attempt to elaborate. Recent progress in statistical concepts and methodology has been amazingly rapid, and certainly the ASQC deserves great credit for the stimulation and education it has provided. My only comment is to caution that we do not get so wrapped up in our mathematical methods that we get like the statistician's son, in a natural science class then studying the nesting habits of birds, who maintained that the domestic hen was unique in laying her eggs on an average. It is not unknown in the history of quality control that other eggs have been laid on an average!

Policy - Directive Control of Quality Excellence

Promotion - Regulative Control of Quality Excellence

To save time I shall discuss these two together. They may seem rather nebulous at first. Yet except for that part of quality control necessary solely to keep processes operable and efficient, all other quality control matters should be within the framework of a definite quality policy. Quality excellence is a product feature in the same sense as price, style, and technical suitability. It is one of the factors which must be weighed in programming both production and sales. It may be of key importance for some companies, for some products, for some markets. Product excellence may amount to very little under other circumstances where competing factors, such as price, dominate.

A research analysis of a company, product, and market situation might, for instance, lead to the conclusion that the corporate fortunes would be favored by orienting the whole operation to a price game,

cutting every corner to get products out at so low a cost that no one
would expect much in the way of quality excellence. In another situa-
tion the odds would favor major attention to style, variety, and versa-
tility at whatever quality and price might result. In a third case,
success might be predicated on outstanding product excellence that would
gain a premium price and insure customer loyalty. It should not be as-
sumed that excellence is a virtue that brings its own reward if to at-
tain it one is at a disadvantage on other factors. These are the con-
siderations referred to in speaking of quality policy.

Analysis of business failures in this industry show many examples
of companies which made a fetish of product excellence, out of tradi-
tion or in faith in the better mousetrap theory, but couldn't survive
in a market where success went to those playing the price or novelty
game. A list of such failures also includes many which had no real
opportunity to compete on a cost or versatility basis, and failed to
seize an available opportunity to build a position on product excellence.

It is my contention that responsibility for determining an appro-
priate quality excellence policy is a proper function of quality control,
in the expanded role I am suggesting. The most skillful quality control
system in the business could gain the company little or in fact be self
defeating if the assumed target is improper for the circumstances. Such
policy itself should be under continued review, for in this day and age
things change rapidly.

I am reminded of a greige mill that found itself in some difficulty
on general quality of an apparel fabric very important to their program.
They had followed the usual tradition of grading similar goods, and had
failed to reckon with changed conditions, in that fashion was calling
for very revealing plain shades and that resin treatments were being
applied. With the combination of these two factors, minor irregulari-
ties and details that were not previously even noted became glaringly
accentuated. After a first reaction that the market was unreasonable
to expect the demanded perfection, the mill finally came around and did
one of the most intensive quality improvement jobs I have ever known.
It was a classic in what a rigorous quality control program can accom-
plish, though, of course, it took time. Finally, after this campaign,
the product was admitted by all as the number one make and in a class
by itself. The mill now experienced no complaints, but also, unfor-
tunately, almost no orders.

In a management meeting to try to figure out this dilemma, a
rather blunt salesman shocked the group by asserting that the only
trouble with this mill was that it didn't make enough seconds. He
was not thanked for his opinion at the time, but later analysis showed
that this mill's seconds and competitors' goods that were no better,
were taking over the current market at the prices at which they could
be sold. Fashion now called for bizarre and eye-catching prints, and
flowing and heavily detailed garment design, which completely obscured
the types of irregularities and defects in the greige goods that the
mill had labored so hard and expensively to overcome. Why pay for a
quality excellence that no one would see?

By hindsight, the lessons in this tale are quite apparent. The
real point of the story is that there were ample facts, available
sufficiently in advance of the two predicaments in which this mill

found itself, which would have guided quality policy so as to capitalize on, rather than suffer from, both situations. Quality control must, if it is to fulfill its full role, keep one eye on the future as well on the problems of the day. Part of this assignment is in a specialized type of market research which is only beginning to be recognized, which determines what part product quality will play in winning a favorable position. Part is in internal research to assess quality capabilities of the plant and its organization. The summation is a type of operations research calculation in which all factors, of which quality is only one, are balanced out for optimum benefit.

Once policy is set and provision made to keep it current, there is the job of implementing it along previously mentioned lines. In addition, there is the continuous watchdog responsibility to see that operational decisions are consistent with the policy. Aggressively, this quality promotion is in pointing out opportunities in equipment, work loads, supervision, training, maintenance, etc., which would favor attainment of quality goals. Defensively, it is in seeing to it that decisions made primarily for cost or other reasons do not prejudice ability to achieve desired quality. All too often, the quality control staff has to struggle with a situation it had no part in creating.

Summary

These comments hit only a few high spots of what is a complex subject. My intent is only to suggest that perhaps the methodology of quality control has advanced further than skills in applying quality control for maximum benefit. In the textile business we have not only to contend with rough technical control problems due to the nature of our materials and processes, but also have the handicap that typical textile organizations have yet much to learn in coordinating their staff functions to full advantage. There are opportunities within the broad meaning of "Quality Control", however, to both improve quality in a relative sense, and to improve the use of quality character- ter and quality excellence as potent features of corporate programming. This expanded role of quality control endeavor must draw heavily on the attitudes and techniques which by definition are true research principles. There is reason to believe that some of our routine quality control procedures will prove unnecessary if we can probe deeper into our problems. Let us at least be sure we do not lose sight of "quality" by over-obsession with quantitative factors.

ADMINISTRATIVE APPLICATIONS: SOME COMMENTS AND SUGGESTIONS*

Lloyd A. Knowler, Chairman
Department of Mathematics and Astronomy
State University of Iowa

The primary purpose of this report is to call atten-
tion to some of the elementary principles of Quality Con-
trol by Statistical Methods or Operations Research which
are being applied to administrative procedures and possibly
to suggest some additional approaches.

In administration, as in manufacturing or any other
division of an organization, management is concerned with
the production of a product or service at the least cost
per usable item. Consequently it is necessary to evaluate
the work in each division in relation to the overall opera-
tions and to use such methods and techniques as will con-
tribute to the objectives of the organization. In so
doing it is desirable to have the best arrangement of
functions and duties involved in the operations. To con-
tinue toward the achievement of this objective will pro-
bably require at least a periodic study of existing condi-
tions and possibly a gradual change of operations or a
shift of responsibilities. A quality control engineer
is particularly trained to make such studies and he must
be used in this capacity if he is to make his greatest
contribution to the organization.

A specific request was made to include "clerical
application of statistical methods" in this report--one
of the divisions referred to above.

A brief survey of the work in any office indicates
the presence of a great deal of administrative detail
including clerical work. Undoubtedly most, if not all,
of the detail is, or has been, desirable. Even though
some of the chores could be eliminated, there is often
some hesitancy if not reluctance to dispense with them.
It is observed that the installation and improvement of
mechanical apparatus eliminates many of detailed operations
and makes work much more pleasant. However, because of
the additional information which can be supplied by the
apparatus and because of the additional demands made upon
it we find that a comparable elimination of administrative
detail is not always evident. In fact, the opposite
often prevails.

One of the things which would concern the quality
control engineer would be an evaluation of the needs for
the various operations being performed or being contem-
plated. Naturally such evaluation would be made in con-
sultation with management. One executive remarked that
his organization spent several thousand dollars to learn
how to do one phase of their work rapidly and then they
discovered that they could abolish the whole operation
with the use of a quality control chart.

Among the specific examples discussed were the
following:

1. The application of statistical methods to the work
   of a pool of typists whose work is somewhat varied
   but at the same time somewhat routine. The items of
   error hazards, costs and production were discussed
   in connection with the use of control charts;

2. Extensions of the principles illustrated in example
   1 were made to other situations such as transporta-
   tion, clerks, sales and printing;

3. The application of statistical methods to a small
   operation involving three or four persons or even
   one person; and

4. The application of statistical methods to duplicate
   readings such as laboratory tests.

PROCEDURES AND TABLES BASED ON
ACCEPTABLE RELIABILITY LEVELS

D. A. Hill and R. H. Myers
Hughes Aircraft Company

Abstract

Quantitative tests for reliability of equipment were started in 1955. The method was developed into procedures and tables applicable to many types of parts and equipments and over a wide range of Acceptable Reliability Levels. The paper includes experience in applying this approach to major supplier black boxes of complex systems.

Introduction

Electronics has come into its own in the period since World War II. Every type of offensive or defensive tactic depends on black boxes, whether they chart a path in space for missiles, provide eyes and ears for submarines, or link the operations of widely scattered ground forces. Industry has turned to electronics to handle its great quantities of data and increasingly to control its processes and machines. The six tube receiver has become a two to twenty thousand tube "brain" with capacity for memory, calculation and programmed decisions. This enormous complexity, together with the unfriendly environments in which delicate equipment must live, brought about the problem of reliability. The question affects other types of equipment, but electronics has had the most difficult job. The effort to find solutions has concentrated in this field.

When man seeks control over a problem, he looks for something he can measure. It is the essence of quality control that the indefinite characteristics of products are replaced with quantitative measures. We want to make predictions, watch trends and evaluate the result of changes. Reliability presents a different problem from other equipment parameters. It necessarily includes an added dimension -- time. This is generally taken for granted when we consider length, weight, voltage output or chemical content. When we test for reliability, we may sample a percent of available equipments but we are also sampling time. In addition to our interest in reliability trends and control, we are also trying to make a reasonable prediction of what will happen during subsequent field use.

Objectives

Important techniques[1],[2],[3] contributing to designing acceptable reliability levels into complex electronic systems have been formalized and are being applied by a few companies. Regardless of how valuable these design procedures become, reliability testing will still be mandatory for purposes of (1) detecting and correcting unforeseeable

Figure 1. Factory Test System Failure Pattern Showing an Essentially Constant Failure Rate Following System Harmonization and Debugging.
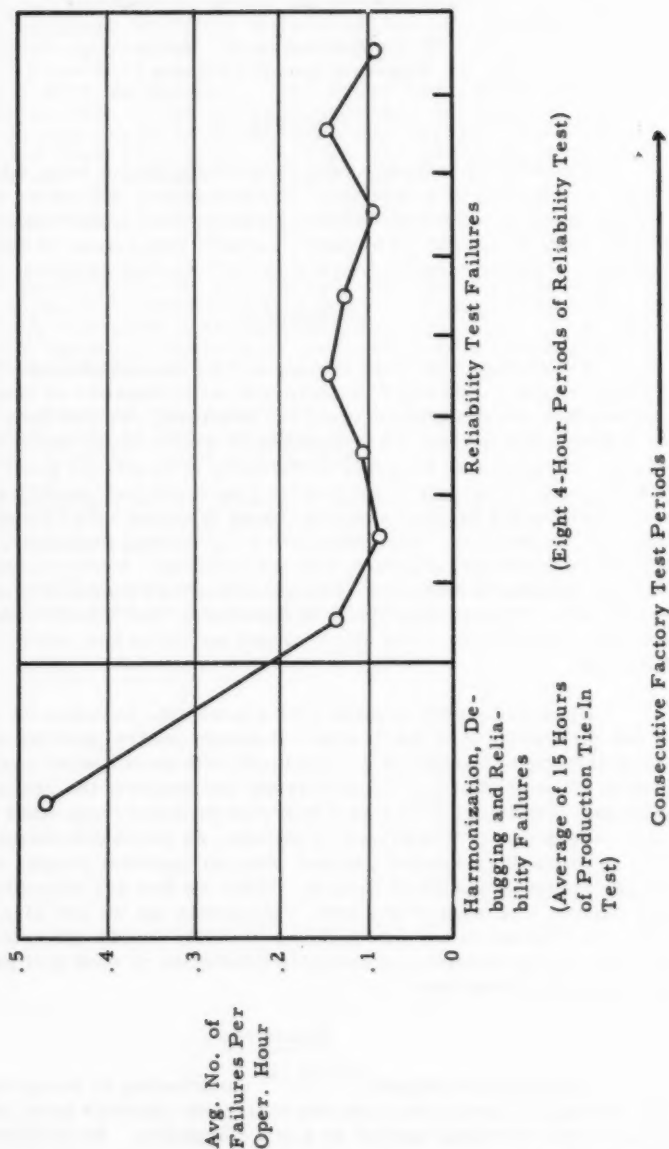
564

part, unit, or system modes of failure, (2) verifying and revising parts failure rates used to estimate reliability from the drawing board, (3) determining hardware conformance to specified reliability levels, and (4) monitoring reliability trends and evaluating the results of changes.

It is the latter two areas with which this paper concerns itself; presented are technical, economic, and statistical aspects and experience relative to quantitative reliability testing procedures.

For tests on manufactured equipment, the objective was to develop a technique that
    —is practical to run in a manufacturing operation
    —will provide a quantitative measure of equipment reliability
         before shipment
    —can be correlated with performance results in the field
    —helps to highlight the problems that need action.

With the establishment of overall requirements on the systems being supplied to the services, it became important that methods be developed for establishing reliability control at system, black box and part levels.

The test methods described in this paper have been applied principally to major systems, such as weapons control, navigation and communication equipment, and to the subsystems and black boxes that go into them. The detailed parts like transformers, relays and gyroscopes have generally been covered by the application of environmental and life test requirements.

In line with the above objectives, a tabulation of reliability testing procedures and tables were developed. They comprise detailed instructions and standardized tables from which time sampling plans may be selected and incorporated into reliability testing specifications. Specific development and application considerations related to Acceptable Reliability Level techniques are described in succeeding sections of the paper.

Reliability Testing Procedures and Tables

The time sampling plans derived were mathematically based (see the Technical Appendix for related definitions and formulae) on a failure distribution per operating time in a chance or random manner describable by a Poisson process, i. e., an essentially constant failure rate. Various field studies by ARINC[4] and Vitro[5] had confirmed that the Poisson process describes a large percentage of complex equipment failure patterns. Comparable Hughes[6] results based on factory test data are shown in Figure 1. Each point represents the average number of failures per operating hour of test (for a large group of systems) over initial debugging and harmonization, and subsequent reliability test periods.

565

Figure 2. A Portion of the Master Table for Normal Reliability Testing Covering the Range of 1.5 to 250 Failures per 1000 Operating Hours.

TABLE III-A

MASTER TABLE FOR NORMAL AND TIGHTENED RELIABILITY TESTING
(SINGLE SAMPLING)

Sample Test Time Code Letter	Sample Test Time	Acceptable Reliability Levels (ARL's) — Normal Testing											
		1.5 Co No	2.5 Co No	4.0 Co No	6.5 Co No	10.0 Co No	15.0 Co No	25.0 Co No	40.0 Co No	65.0 Co No	100.0 Co No	150.0 Co No	250.0 Co No
A	75	↓ 0 →	0 1	→		1 2	2 3	3	5 6	8 9	12 13	18 19	28 29
B	100	0 ↓←	↓		1 2	2	3 4	5	7 8	10 11	15 16	23 24	36 37
C	150	1 ←	↑	1	2 3	3	4 5	7	10 11	15 16	22 23	33 34	51 52
D	250	1 2	1 2	2 3	3	5 6	7 8	11 12	16 17	24 25	35 36	51 52	79 80
E	350	2	2 3	3 4	5	7 8	10 11	15 16	22 23	33 34	48 49	69 70	←
F	500	2 3	3	4 5	6 7	9 10	13 14	20 21	29 30	45 46	68 69	←	
G	750	3 4	4 5	6 7	9 10	13 14	19 20	29 30	43 44	66 67	←		
H	1,150	4 5	6 7	8 9	13 14	19 20	26 27	40 41	60 61	←			
I	1,500	5 6	8 9	11 12	17 18	24 25	34 35	54 55	←				
J	2,200	8 9	11 12	17 18	24 25	34 35	49 50	←					
K	3,000	10 11	14 15	20 21	32 33	44 45	←						
L	4,500	14 15	20 21	28 29	43 44	61 62							
M	7,500	20 21	31 32	44 45	←	←							
N	15,000	←	←	←									
O	30,000												250.0 ✕
P	50,000	2.5	4.0	6.5	10.0	15.0	25.0	40.0	65.0	100.0	150.0	250.0	

Acceptable Reliability Levels (ARL's) — Tightened Testing

↓ = Use first sampling plan below arrow.

↑ = Use first sampling plan above arrow.

Sample Test Times are in terms of operating hours.

Acceptable Reliability Levels (ARL's) are in terms of failures per thousand operating hours.

The failure rate declines rapidly as debugging, adjustment and harmonization proceeds. This is followed by a relatively constant and random failure pattern.

To facilitate application sampling plans were tabulated. As shown in Figure 2, all time sampling callouts are in familiar engineering terms of operating hours and related failures. The Acceptable Reliability Level, expressed as number of failures per thousand operating hours, was termed the ARL. By using the index, failures per thousand operating hours, the range covered by the tables can be most conveniently expressed numerically. Failures per thousand operating hours are additive. This permits us to determine system reliability from the sum of test results on subsystems (when independence is assumed). The reciprocal of ARL, or Mean Time Between Failure, cannot be added. In addition most parts failure data is given in terms of failures per some base of operating hours, so that studies of the relation between parts and systems are facilitated.

The range of available plans extends from 250 to .001 failures per thousand operating hours. 250 failures per thousand operating hours (equivalent to a mean time between failure of 4 hours) is a lower limit that might be applicable to a very complex system. Conversely, .001 failures per thousand hours is an appropriate requirement for a single part such as a capacitor.

As shown in the Figure 3 example, matched sets of single, double and multiple time sampling plans are available which have essentially the same risks in their ability to evaluate product conformance to the specified reliability level. Choice of which type of reliability testing to use can, therefore, be based entirely on economic versus administrative considerations. For example, single sampling administrative advantages can outweigh multiple sampling test time savings in reliability tests of complex systems.

Provisions for tightened and reduced testing have also been provided. Sample Operating Characteristic curves for Normal, Tightened and Reduced Time Sampling Plans are outlined in Figure 4. The tightened reliability criteria (less allowed failures) are instituted to protect consumer interests when prior test results reflect an unacceptable reliability level requiring product improvement. Conversely, reductions in test times or frequency are permitted (with related decreased costs) when acceptable reliability levels have been demonstrated and sustained.

Although the procedures were developed primarily for purposes of production reliability control, specific time sampling plans (since they are independent of the number of items tested) are applicable for use by Development groups in qualification type reliability testing where few equipments are available. In this Engineering area, use of tightened conformance criteria is recommended so as to minimize the

Figure 3. Matched, Single, Double or Multiple Sampling Plans For
Acceptable Reliability Level (ARL) 4.0.

ARL	Type of Sampling		Sample Test Time (hrs.)	Cumulative Test Time (hrs.)	Failures Observed	
					Conforming	Non-Conforming
4.0	Single		1500	1500	11	12
4.0	Double	1st	1000	1000	7	17
		2nd	2000	3000	16	17
4.0	Multiple	1st	400	400	1	6
		2nd	400	800	4	8
		3rd	400	1200	7	11
		4th	400	1600	10	14
		5th	400	2000	13	17
		6th	400	2400	16	21
		7th	400	2800	20	21

568

chances that product which passes a development reliability test, will fail the production counterpart due to sampling fluctuations.

The time sampling plans and related procedures were derived along similar lines to MIL-STD-105A, considered representative of conventional inspection sampling methods. The military standard establishes a sample size or number of pieces to be drawn from a lot presented for acceptance and gives the acceptance number of defectives allowed in this sample. The reliability testing tables designate a sample of operating hours to be observed on one or more units of product and give conformance number of failures. Where MIL-STD-105A refers to Acceptable Quality Level (AQL) these tables use the term Acceptable Reliability Level (ARL) to denote the minimum acceptable levels of quality and reliability respectively.
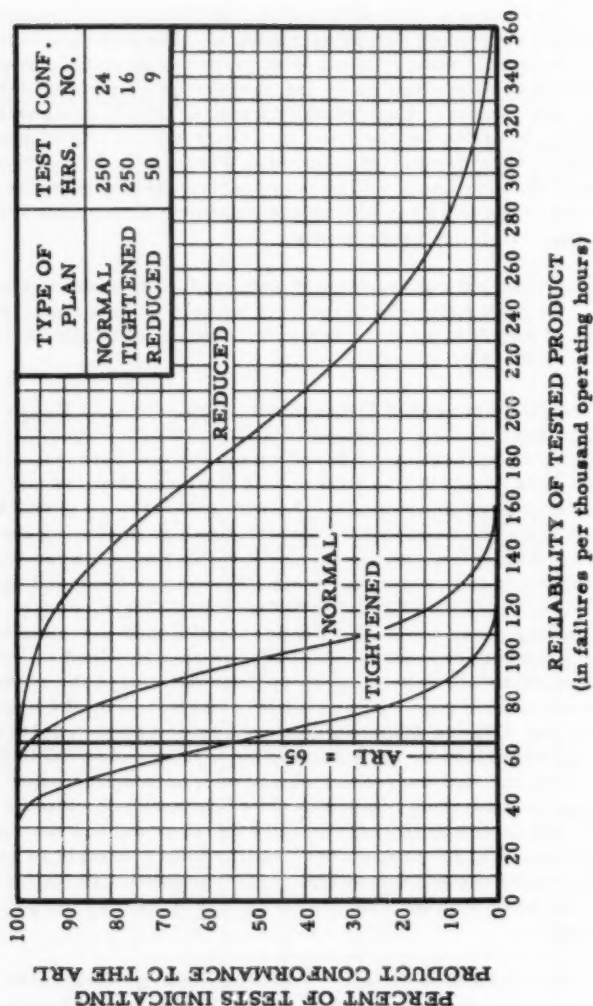
The concepts of consumer* and producer* risks are also quite similar. In each case, the larger the sample or test time, the lower the risks of considering as acceptable, product of quality or reliability poorer than the AQL or ARL.

The Department of Defense set up an Advisory Group on the Reliability of Electronic Equipment (AGREE). Their report[7] includes an approach to acceptance tests based on mean time between failure and a comparison is of interest. Figure 5 shows ground rules and considerations relative to each approach. As outlined both are predicated mathematically on a Poisson failure distribution per operating time. The AGREE reliability index is in terms of mean time between failure, the HAC Tables use the reciprocal term failures per thousand operating hours. HAC procedures provide for matched single, double or multiple time sampling plans while AGREE has established a sequential reliability testing procedure. Where AGREE has standardized on various levels of test environments depending on operational use of the equipment, HAC procedures consider that these will be established by the applicable test specification. Where demonstrated nonconformance requires continuation of, or reversion to 100% reliability testing under AGREE, HAC ground rules call for corrective action and/or suspension of shipments in addition to instituting a more severe conformance criteria (tightened testing). When reliability test results equal or exceed the specified reliability level AGREE procedures require only a sample of equipments to be reliability tested; similarly, HAC procedures permit a reduction in total test time and/or a reduction in test frequency, e.g., monthly to bi-monthly.

Test risks also differ. Under AGREE ground rules producer and consumer risks were fixed at about 10% for an operating ratio* of 1.5 to 1. Consistent with the above risks, AGREE tests require, if equipment is at the specified mean time between failure (MTBF), about 20 times the specified MTBF to reach a decision. Conversely, HAC

*Defined in Section 1.0 of the Technical Appendix.

569

Figure 4. Operating Characteristic Curves for Normal, Tightened and Reduced Time Sampling Plans for Acceptable Reliability Level (ARL) 65.

TYPE OF PLAN	TEST HRS.	CONF. NO.
NORMAL	250	24
TIGHTENED	250	16
REDUCED	50	9

RELIABILITY OF TESTED PRODUCT
(in failures per thousand operating hours)

PERCENT OF TESTS INDICATING
PRODUCT CONFORMANCE TO THE ARL

producer, consumer risks and operating ratios are variables with related test times preselected on the basis of test equipment and manpower availability, associated costs and the risks the buyer is willing to live with. For example, less complex, more reliable black boxes will require a longer test for an adequate measure of their reliability. On the other hand, these boxes will be subject to further tests as a part of the system. It may make sense to economize at this point and accept a larger risk. Subsequent tests will catch up with unfavorable trends. The completed system will have a much higher failure rate than any of its component items. A relatively shorter time-sample will do the job.

Application to Reliability Specifications

With overall requirements established for the systems being supplied to the military, it was necessary to pass on to major black box suppliers compatible reliability criteria in the form of routine, consistent and economic test specifications.

Routine specifications were required so that the inherently lengthy cycle of test program development, supplier negotiation, reliability test, evaluation of results, corrective action, retest, etc., would be minimized.

Consistent specifications were necessary so that the test criteria by which purchased black boxes were judged, and which control test equipment and test operating costs, were equitable, economic and specific. In addition, it was advantageous to have available consistent material for use as guides to suppliers in analogous application to their own operation.

Economic specifications were needed so that black box reliability levels were evaluated at minimum time and cost through routine application of test and sampling procedures.

First, a standardized military type specification format was established. Called out were the following nine (9) paragraphs. Considerations appropriate to each are described below.

1.0 Purpose

Contains a brief statement of the specification objectives. For example, "To outline the requirements of a reliability measurement program designed to determine conformance to the specified reliability level. "

2.0 Scope

Contains a tabulation of black boxes and accessories to which the reliability test is collectively applicable. In this area considerable savings in time and costs are possible through maxi-

Figure 5. Comparison of Ground Rules Used in Deriving AGREE and HAC Reliability Testing Criteria.

Ground Rules	AGREE Production Testing	HAC Procedures & Tables
Mathematical Basis	Poisson Distribution	Poisson Distribution
Index of Reliability	Mean Time Between Failure	Failures per 1000 Operating Hrs.
Type(s) of Time Sampling Plans	Sequential	Single, Double, or Multiple
Test Conditions	Standardized	Considered part of the Applicable Test Specification
Test Risks	Fixed	Variable
Demonstrated Nonconformance	Revert from Sampling to 100% Test	Corrective Action and Revert to Tightened Testing
Sustained Conformance	Sampling Permitted	Reduced Testing Permitted
Test Times	Averages 20 times MTBF	Varies from 05 to 60 times MTBF

imum group testing of black boxes by subsystems and suppliers.
On the other hand, test equipment size or schedule restrictions
may dictate that a complex system be subdivided into several
functional reliability test loops.

3.0   Applicable Specifications and Standards

Appropriate statements as to what specifications or standards
form a part of the reliability test specification. For example,
applicable product specification, test procedure, sampling pro-
cedures and tables, etc. Confusion will result if the buyer and
supplier do not agree on a complete description of the test con-
ditions and criteria.

4.0   Acceptable Reliability Level

Contains a quantitative description of the minimum acceptable re-
liability level in terms of both failures per thousand operating
hours and equivalent mean time between failure. Since the speci-
fied reliability level and related conformance requirements domi-
nate the reliability test specification, more detailed discussion re-
garding the development of the reliability requirement is in order.

Acceptable reliability levels established for black boxes must be
consistent with overall system requirements. Specifically this
means all black box failure rates must add to the system failure rate.
Where black boxes are common to more than one system, it may
be necessary to set the reliability requirement at a level that will
meet the most demanding application. Once derived (on a similar
black box and/or complexity basis) reliability requirements per-
mit of effective implementation in the following related areas.

Black box failure rates constitute a logical priority base for gen-
erating test specifications, on the basis of highest failure rate
items first. Concentrated effort on the highest failure rate equip-
ments would seem to be the most rapid and economical means of
reaching the system requirement. What can be gained by con-
centration on black box testing instead of waiting to test a com-
plete system? First, calendar time savings, since reliability
testing can be instituted and improvements effected much earlier
in the development, production, field test cycle. Second, it is
possible to generate a relatively larger number of test operating
hours resulting in more failures; therefore, related corrective
action can be instituted earlier. Third, a more intensive relia-
bility evaluation is usually possible due to more detailed test
procedures, test instrumentation or test conditions.   Also of
value is that black box requirements when compared with relia-
bility test results spotlight those units farthest from predicted
values on which the most fruitful reliability improvement effort
can be invested.

573

5.0   Definitions

Contains standard definitions of the primary phrases and state-
ments found in the various paragraphs; for example, reliability,
index of product reliability, failure, fault, mean life, acceptable
reliability level, etc.

6.0   Conditions of Reliability Test

This paragraph covers the following primary considerations:
Establishes a standard test operating cycle;
Establishes the specific test procedure to be used to deter-
    mine malfunctions;
Establishes the test conditions (i.e., ambient, environ-
    mental or combinations thereof) under which testing is
    conducted.

As outlined above, the standard test cycle attempts to provide a
meaningful simulation of actual operating conditions. Specific-
ally, it establishes when the equipment shall be turned on, exer-
cised in various tactical modes, checked for malfunctions
against the test procedure, turned off, etc.

Repetition of the standard test cycle achieves the total operating
time required by the applicable time sampling plan callout. This
makes for maximum flexibility since the time sample can be in-
creased or decreased, and can use fixed or sequential type sam-
pling, without affecting the test procedure or standard test cycle
requirements.

In addition to the standard test operating cycle described, para-
graph 6.0 should clearly establish the reliability test procedure
to be used to determine malfunctions. This should be based on
an existing test procedure, test specification, product specifica-
tion, etc. Regardless of its base the reliability test procedure
should have the following characteristics: be specific; be avail-
able or readily generated; and (appropriate to black boxes) equal
or exceed in severity the reliability test requirements the boxes
must later pass at the systems test level. In this sense, in-
creased severity can mean more characteristics are checked,
tighter parameter tolerances are called out, or the application
of realistic environmental stresses.

To summarize the approach to test design: the performance test
procedure (reliability is essentially performance over time), in
conjunction with a repeated standard test cycle, under ambient
and/or environmental test conditions provides an economic,
meaningful and rapidly effective reliability test.

574

Figure 6. Subsystem A Mean Time Between Failure Reliability Test History.
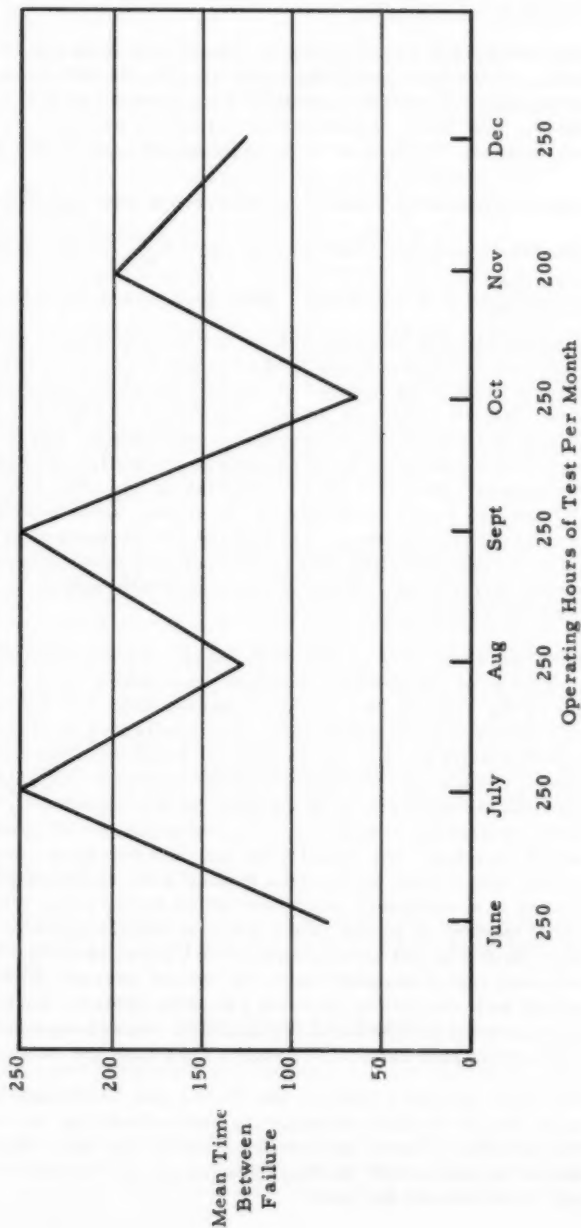
7.0   Terms and Conditions

This paragraph comprises methods of sampling and reporting inputs.   Methods of sampling together with the test conditions of paragraph 6.0 jointly control test equipment and test operating costs.   The latter is predicated primarily on engineering considerations, the former on statistical and probability principles.

The recognition of mean time between failure ($\frac{Time}{Failures}$) as a statistical variable (like yield strength ($\frac{Force}{Area}$) but unlike current ($\frac{Voltage}{Resistance}$)) is important.   What this means is that the calculated mean time between failure (or failure rate) is subject to sampling fluctuations and tends to vary from the "true" value more widely if the number of hours used in computing the mean time between failure is small.   As shown in Figure 6, this variability is reflected in Subsystem A experience, where seven (7) successive operating hour periods of reliability test indicated subsystem mean time between failure of 83, 250, 125, 250, 62, 200 and 125 hours respectively. In effect, since reliability test results inherently vary, it is logical that statistical criteria, such as time sampling plans, allowing for such fluctuations, be used to determine conformance or nonconformance to the specified reliability level.

Sampling flexibility is also necessary.   Procedures should be independent of the number of equipments tested, thereby permitting use in the critical development phases, when few items are available for reliability evaluation. The inevitable question of whether to test 4 units for 200 hours each or 8 units for 100 hours also arises in this area. For all practical purposes, the answer to this question would seem to be dictated by test equipment, manpower, time, or product availability. For example, development qualification type tests will usually be run on a few items for many hours; conversely, production testing will comprise periodic testing of a relatively large number of boxes for a relatively small number of hours. This inherent development-production relationship is not undesirable since it permits early evaluation of wear-out type component failures, whose average times to failure extend well out into the service life of the system. Such failures are amenable to scheduled replacement and subsequently considered a much less critical reliability area than the random failure.

The other primary input in the Terms and Conditions paragraph concerns the method of reporting and submitting the reliability test results.   These indicate what reliability test information should be submitted, backup data required, frequency of reporting, to whom and by when.

576

8.0   Corrective Action

Of particular interest to both supplier and buyer is what happens when nonconformance to the reliability requirement has been demonstrated.   Several approaches are available.

Nonconformance can mean:
>100% testing of product in lieu of sampling, or
>sampling is continued, but corrective action and/or suspension of shipments is required.

Representative of the first approach are AGREE production reliability testing procedures.   There is no doubt that the 100% test following nonconformance provides the supplier with considerable corrective action incentive in order to preclude extensive schedule slippage and increased test costs. However, where reliability is at an unacceptable level the necessary corrective action, more often than not, requires finding better parts, redesigning circuits, or improving manufacturing processes.   This takes time, and 100% testing contributes little to the solution of known problems.

The alternative (continued sampling but concurrent corrective action) also has pros and cons.   Although providing less incentive, a supplier would be more inclined to accept a higher (but achievable) production mean time between failure on a corrective action basis in lieu of a lower reliability level when production schedules are involved.   In electronics the state of art is such that it is better to shoot for a goal that is just out of reach, rather than fix a level that is achievable but inadequate.   Reliability testing on a fixed test time basis also has the economic advantage of precluding investment in standby test equipment and manpower necessary to offset possible 100% test consequences.

9.0   Preparation for Delivery

Contains a brief standard statement covering costs, disposition and repair of failed units of product which have been reliability tested.

Conclusions and Results

The more important conclusions and results were considered to be as follows:

With system reliability requirements and tests becoming standard contractual practice, it became important that rapid, consistent, economic means be developed for establishing reliability control for subsystems and black boxes.

Reliability Testing Procedures and Tables were subsequently derived comprising detailed instructions and standardized tables from which time sampling plans were selected and incorporated into reliability testing specifications.

Since major supplier black boxes comprise almost 40% of newer systems, application effort has been concentrated in this area, with full recognition that supplier reliability testing must complement, not replace, timely component and system evaluations at other stages of development and production.

Use of a standardized military type format permitted economic development and rapid initiation of quantitative reliability test specifications.

To date reliability testing and related systematic gathering of data has proved to be an effective method of monitoring the results of improvement effort.

Regardless of increased application of design reliability procedures, reliability testing will still be mandatory for purposes of (1) detecting and correcting unforeseeable part, unit, or system modes of failure, (2) verifying and revising parts failure rates used to estimate reliability from the drawing board, (3) determining hardware conformance to specified reliability levels, and (4) monitoring reliability trends and evaluating the results of changes. Reliability is in the process of joining the established requirements for system design such as power required, output, sensitivity or weight. In the future, it will be controlled on a quantitative basis.

Technical Appendix

1.0   Definitions and Assumptions

1.1   A discrepancy is any malfunction which occurs during the course of reliability testing when a system, unit, part, ceases to perform satisfactorily. A discrepancy is considered to occur at a discrete point in time which is the instant of observation of the malfunction. For purposes of reliability testing discrepancies have been subdivided into failure or fault categories.

1.2   A failure shall be considered as

1.2.1   Any out-of-adjustment condition other than those which are normally readjusted by a pilot-operator in flight.

1.2.2   Any part failure which is not caused by test equipment malfunction, procedural error, or by failure of another part.

578

1.2.3 Any malfunction whose cause is unknown.

1.3 A fault shall be considered as

    1.3.1 Malfunction or damage due to the test equipment, incorrect procedures, or mishandling.

    1.3.2 Out-of-adjustment conditions which are normally pilot-operator adjusted in flight.

    1.3.3 Secondary failures resulting directly from the failure of another component.

    1.3.4 Manufacturing error, such as use of the incorrect part number, that can be readily corrected in all work in process and which will not recur in service use.

1.4 Operating time comprises the total time one or more sampled units of product are operated in accordance with the applicable reliability test specification.

1.5 The reliability index is expressed in terms of failures per thousand operating hours.

1.6 Failures per thousand operating hours (f) expressed as an equation is

$$f = \frac{F \times 1000}{T} \qquad (1)$$

where F equals total failures per total operating hours T. Failures per thousand operating hours divided into 1000 is the equivalent of mean time between failure (MTBF) or mean life (ML). Expressed as an equation, mean time between failure (MTBF) or mean life (ML) is

$$MTBF = ML = \frac{1000}{f} = \frac{T}{F} \qquad (2)$$

where f, T and F are as defined above.

1.7 Producer risk is the probability that product at the acceptable reliability level (ARL), when reliability tested will be considered nonconforming, due to sampling fluctuations.

1.8 Consumer risk is the probability that product at a specific reliability level poorer than the ARL when reliability tested will be considered conforming due to sampling fluctuations.

579

1.9   Operating ratio is the ratio of the Acceptable Reliability
       Level in terms of mean time between failure (or failure
       rate) to a poorer reliability level at a specific consumer
       risk.

2.0   Mathematical Basis

2.1   The time sampling plans derived are predicated on a
       failure distribution per operating time in a manner de-
       scribable by the Poisson process, i.e., an essentially
       constant failure rate.

3.0   Related Formulas

3.1   Based on the definitions and assumptions of paragraphs
       1.0 and 2.1 the applicable Poisson distribution may be
       expressed as follows:

$$P_c = \frac{(T/ML)^c e^{-(T/ML)}}{c!} \tag{3}$$

where P_c is the probability of c failures occurring per
total operating time T, ML is the mean life or average
operating time between failures, and e is the exponential
constant (2.7183).

When deriving the probability of zero (0) failures during
total operating hours T, formula (3) for $c = 0$ reduces to

$$P_0 = e^{-(T/ML)} \tag{4}$$

usually known as the exponential distribution.

References

(1)   "Reliability Stress Analysis for Electronic Equipments," RCA
       Technical Report TR 1100, October 1, 1956.

(2)   "Designing for Reliability," N. H. Taylor, Lincoln Lab, MIT,
       Proceedings of the IRE, June 1957.

(3)   "Reliability of Systems," O. B. Moan, Lockheed Aircraft
       Corp., Control Engineering Data Files 2, 4, 5, February,
       April, May 1957.

(4)   "Air Force Reliability Assurance Program Progress Report
       No. 1," Aeronautical Radio, Inc., February 15, 1956.

(5)   "Techniques for Reliability Measurement and Prediction Based on Field Failure Data, " Vitro Laboratories, Technical Report No. 80, October 10, 1955.

(6)   "Time Samples Measure Equipment Performance, " D. A. Hill, R. H. Myers, and H. D. Voegtlen, Hughes Aircraft Co. , Report R-56-7, October 10, 1956.

(7)   "Reliability of Military Electronic Equipment, " Advisory Group on Reliability of Electronic Equipment, Dept. of Defense, June 4, 1957.

RELIABILITY - A ROLE FOR QUALITY

A. L. Lambert
The Martin Company
Baltimore, Maryland

I. Introduction

A critical attitude, a high degree of curiosity and an abi-
lity to evaluate objectively - these characteristics are valuable
tools to those who work in Quality. These tools, useful in every-
day work, should be helpful in determining the responsibilities of
Quality in dealing with reliability.

The need for curious, critical examination and objective
evaluation stems, in part, from the uncertainty that cloaks the
reliability field. The nature of the uncertainties are well known.
It is not infrequent that questions are raised with respect to the
importance of reliability, the amount of money that should be spent
for it, how it should be specified and how it should be demonstrat-
ed.

In the face of questions such as these, it might well be con-
cluded that the best Quality approach would be to wait until the
fundamentals are firmly established. In opposition to this are
the need for Quality to make active contributions to the advance-
ment of reliability, and the fact that reliability is now being
specified contractually. Some recent military proposal require-
ments have included not only quantitative reliability specifica-
tions, but also requests for detailed reliability plans with
corresponding work schedules and manpower allocations.

The responsibilities of Quality and its role in reliability
have become matters of immediate significance.

II. A Matter of Viewpoint

The first step in a critical look at reliability is to decide
what it is. One way of doing this is to examine the generally
accepted definition that the reliability of an equipment is the
probability that it will perform a required task under specified
conditions for a required period of time. [1] The definition is
fundamentally important for it permits the numerical specification
of reliability as a probability under defined conditions of per-
formance, use and time.

If we are curious and ask what this definition means, we find,
not unexpectedly, that it means different things when considered
from different viewpoints. To the military it can mean a known
probability of mission success or a maintenance level factor. To
the designer a numerical reliability specification represents a
new design parameter. To Quality the reliability number as a de-
sign parameter becomes a quality characteristic to be measured,
controlled and assured.

This matter of viewpoint would seem to explain some of the
difficulties of understanding and communication that are common
in reliability. Because reliability is a relatively new field, it

583

is easy to seize upon a single viewpoint and exclude examination
of the others; or to gain the impression that reliability is a
hopeless confusion of not too meaningful terms. Actually, of
course, there is definite meaning from each viewpoint and there
need not be inconsistency in saying that reliability is a proba-
bility of mission success - or a maintenance factor - or a design
parameter - or a quality characteristic - or all of these; and
that reliability is a statistical probability under three defined
conditions.

Reliability - Another Grey Area?

From the Quality viewpoint it might be adequate to say that
reliability as a design parameter is a quality characteristic.
It may be helpful, however, to examine the meaning of this
quality characteristic.

There is a tendency to categorize reliability with what might
be called "grey area" product quality characteristics such as
producibility and maintainability (Figure 1(a)), as distin-
guished from characteristics such as performance, weight,
service life and selling price which traditionally have re-
ceived greater emphasis. This approach would result in the
delivery of a product with field characteristics as shown in
Figure 1(b). It is interesting to note that the "grey area"
characteristics generally are those which are not measurable.
Notice that the "grey area" approach to reliability is unsat-
isfactory is spelled out by the contractual specification of
measurable, numerical reliability requirements.

Reliability - A Figure of Operational Quality

The demands of the reliability concept are not satisfied by
the "grey area" approach; but a moments reflection indicates
that even removal of the "grey" from the reliability block of
Figure 2 does not result in an adequate description of the
characteristic.

Reliability as a quality characteristic is more accurately
portrayed as a numerical Figure of Operational Quality which
encompasses the field characteristics of the product. This
concept (or viewpoint) is indicated by Figure 2 which shows,
only in part, the field characteristics of the product.

To consider reliability as a figure of operational quality
is, of course, a matter of relationship and does not remove the
validity of the reliability definition nor the need to treat re-
liability as a design parameter. It is intended to indicate the
proper importance of reliability to Quality and to suggest the
scope of work with which Quality is concerned if reliability is to
be measured, controlled and assured as a quality characteristic.

III. Elements of the Task

To consider reliability as a figure of operational quality
has advantages in terms of perspective. The questions of the re-
sponsibilities of Quality and the role it should play in a

reliability program require, however, an examination of the work to be done. It is then reasonable to consider Quality functions in terms of relationship to the functions of Engineering, Manufacturing, Procurement and other segments of organization.

Reliability requires treatment through all phases of product life from the specification of requirements in the request for bid, through the proposal phase, design, development, production and into service usage. Perhaps the principles of the role of Quality can best be demonstrated by examination of the elements of work in the early phases where Quality activity is usually at a minimum.

Areas for Quality Emphasis

1. Proposal Phase

The reliability task in the proposal phase is shown in brief form in Figure 3. The Request for Bid should reasonably be expected to contain the operational requirements, including the factors of usage, maintenance, and logistics in addition to the numerical reliability requirements. Elements of the reliability task include reliability studies, an estimate of environment, estimates of reliability, preparation of a reliability program and an estimate of reliability costs.

What are the functional responsibilities of Quality in a proposal phase that contains these tasks? The answer, in practice, will depend on the functional part that Quality plays in the particular organization. If, for example, the Quality function is generally limited to examination of the physical product, then the Quality task in proposal preparation may be confined to the submission of manpower estimates to satisfy the requirements of the reliability program. On the other hand, when the Quality function is broad, Quality can play an active part in the proposal preparation phase of the reliability program.

If it is the functional responsibility of Quality to assure that control is maintained and to assure the reliability of the product, then the proposal phase of reliability becomes critical in the discharge of this responsibility. This stems from the basic consideration that the reliability proposal not only permits but also limits, in terms of dollars, the effort to be expended in later phases of the program. It is not only reasonable, but also necessary in terms of Quality effectiveness, that the proposed reliability program contain the necessary elements and be adequate to translate the numerical reliability requirement into a proven design characteristic and into a demonstrated quality characteristic of the end product.

Another example of a factor important in this phase is the specification itself. If vagueness exists in the reliability specification or in the specified means of demonstration - both common types of problems in normal quality work - then the proposal phase offers the most effective time for clari-

585

fication and the elimination of these potentially serious
quality problems.

2. Design and Development Phase

Another area where the influence of Quality is historically
low, but where reliability demands attention, is the product
design and development phase. If we examine the elements of
work in a reliability program in this phase we find a situa-
tion which is basically outlined by Figure 4. The program
consists of: (1) an organizational capability, (2) an ade-
quate definition of the task, (3) a planned series of design
and development steps calculated to satisfy the specified
reliability requirements, (4) controls designed to evaluate
progress and assure the achievement of objectives, and (5) an
adequate demonstration of successful attainment of reliabi-
lity objectives.

Here, again, the functions assigned to Quality will vary from
organization to organization. The traditional tasks such as
receiving inspection and manufacturing in-process control re-
main important and necessary even in the production of the
prototypes and test models associated with this phase. The
effectiveness of Quality in assuring the required operational
quality (i.e., reliability) in the end product is greatly en-
hanced, however, by a more extensive Quality participation in
the design and development phase. It is not intended to
imply that design responsibility be removed from the engineer
or to suggest that Engineering not have primary responsibi-
lity for control of design activities. This primary respon-
sibility, however, does not remove the responsibility of
Quality to assure that controls are adequate to satisfy the
end requirements of the product.

With reference to Figure 4, it is reasonable, for example,
that Quality determine that the organizational capability
exists for execution of the reliability program, that the
definition of the task is adequate, that the reliability
plans and controls contain the necessary elements, and that
adequate provision is made for demonstration. Through the
reporting of defects at the time of demonstration, for
example, it is very possible to determine that an element
such as Human Factors was omitted or inadequately specified
in the initial definition of the task. From the quality as-
surance standpoint it is more effective and of far more bene-
fit to the program if such a discrepancy is discovered sever-
al years earlier by an examination of the task definition
during its formation. Similar comments are applicable to the
essential steps in the reliability program and to the reliabi-
lity controls developed for the design and development phase.

In the control and assurance of reliability as a quality
characteristic, Quality faces this basic situation. The time of
reliability demonstration is too late (probably by several years)
for effective quality control and quality assurance efforts. The
reliability level achieved in the design cannot be improved

significantly by traditional Quality efforts in the manufacturing phase. Last, and by far the most important, is the conclusion that Quality can make its most effective contribution by active participation in the reliability program from the time of inception in the proposal preparation phase.

IV. Application of the Principles

In a discussion of principles it is normally desirable to include a concrete example of an application. For this purpose, the steps which have been taken by Martin, Baltimore are offered for review.

The management organization for reliability at Martin, Baltimore is as shown in Figure 5, with the responsibility for coordination placed within the Quality Division. It should be noted that reliability is recognized as presenting tasks to not only Engineering and Quality, but also to Manufacturing, Procurement and Customer Service. It is pertinent to note that in addition to coordination, Quality has the defined functions; (1) to establish, in conjunction with the various activities, the responsibility for performance of reliability functions, (2) to participate in project reliability planning, (3) to aid in the development of techniques for the analysis, prediction, measurement and attainment of reliability, (4) to perform independent analyses of customer reliability requirements, and (5) to evaluate project reliability plans, efforts and results, and provide effectiveness reports to management for control purposes.

The organizational concept is less than a year old and full implementation is progressing but is not yet complete. The steps which have been taken within Quality are those of interest in discussing the role of Quality and are perhaps best described by the functional organization chart of the Quality and Reliability Assurance Department which was formed about six months ago. The department head reports to the Director of Quality and performs the reliability coordination function described in Figure 5. The shaded areas of Figure 6 indicate the departmental activities which make a direct quality contribution to the Company's reliability effort.

The Quality and Reliability Analysis Section serves as the nerve center in the corrective action feedback loop of the Company's product failure and trouble analysis system. The section collects both field and in-plant information (including information from design and development projects). Beyond the statistical analysis of data, the section performs engineering investigation and evaluation of troubles and failures to select significant problems. Corrective action feedback is provided and is followed by participation of the section in further investigation and the resolution of problems. Another function of the section is that of preparing periodic management reports on product trends and corrective action status.

The quality audit function in the Quality Engineering Section includes the audit and evaluation of reliability programs. As a matter of interest, the basic function calls for audits of

those factors which affect end product quality. This encompasses not only reliability programs and the quality control system, but also functions within Procurement, Manufacturing and Engineering.

The Reliability Engineering activity includes that of assignment of reliability engineers to advance design activities where reliability proposals are prepared. As shown in Figure 6, reliability engineers are assigned to projects having reliability programs in later phases; these engineers participate in the project reliability planning and control efforts.

In this review it is not intended to create the impression that reliability functions are solely a Quality responsibility. This is not the case. It should be noted, in this respect, that the Reliability Engineers assigned to projects are administratively responsible to the projects and functionally responsible to Quality. The prime responsibility for leadership, direction and control of project reliability efforts rests with the Project Engineering Department of the Engineering Division.

The Martin, Baltimore organization for reliability and the selection of Quality efforts are the result of careful study and evaluation of the reliability problem. Experience to date with this approach is felt to be good and is felt to demonstrate the soundness of the approach.

V. Conclusions

The role of Quality in the field of reliability may be controversial for some time to come. Nevertheless, the requirement that reliability be controlled and assured is now a matter of contractual specification. If it is logical that reliability as a design parameter becomes a quality characteristic, then the Quality problem of adequate treatment of this characteristic is real and immediate.

The role of Quality in the treatment of the reliability characteristic may be expected to vary from organization to organization. It is suggested, however, that the most effective role for Quality requires not only the application of traditional efforts but also the acquisition of new skills and the active participation of Quality in the early phases of product inception and planning.

References

(1) Reliability of Military Electronic Equipment - Report by Advisory Group on Reliability of Electronic Equipment, OASD (R & E), 4 June 1957.

588

RELIABILITY - ANOTHER GREY AREA?

Typical Design Parameters

Performance	Weight	Selling Price	Service Life	Producibility
				Maintainability
				Reliability

FIG. 1(a)

Typical Field Characteristics

Performance	Weight	Service Life	Operational Suitability	Maintainability
				Reliability

FIG. 1(b)

RELIABILITY - FIGURE OF OPERATIONAL QUALITY
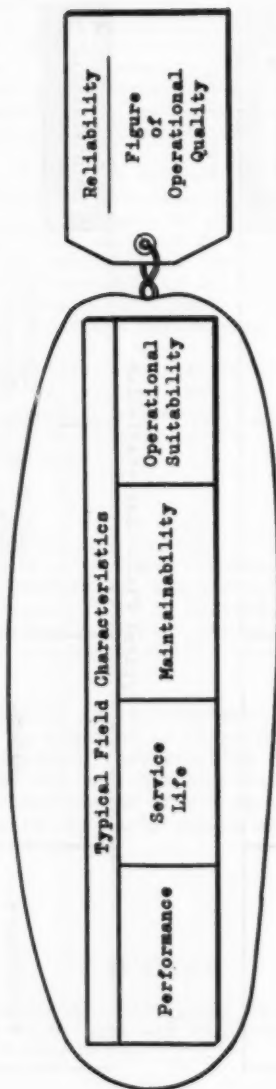
FIG. 2

THE RELIABILITY TASK
PROPOSAL PHASE

Operational Requirements

Usage

Maintenance

Logistics

Numerical Reliability Req'ts.

REQUEST
FOR
BID

Advanced
Design

PROPOSAL
PREPARATION

Reliability Studies

Estimate of Environments

Reliability Estimates

By System & Sub-System

By Operational Req't.

Reliability Program

Reliability Costs

PROPOSAL
TASK
OUTPUTS

FIG. 3

591
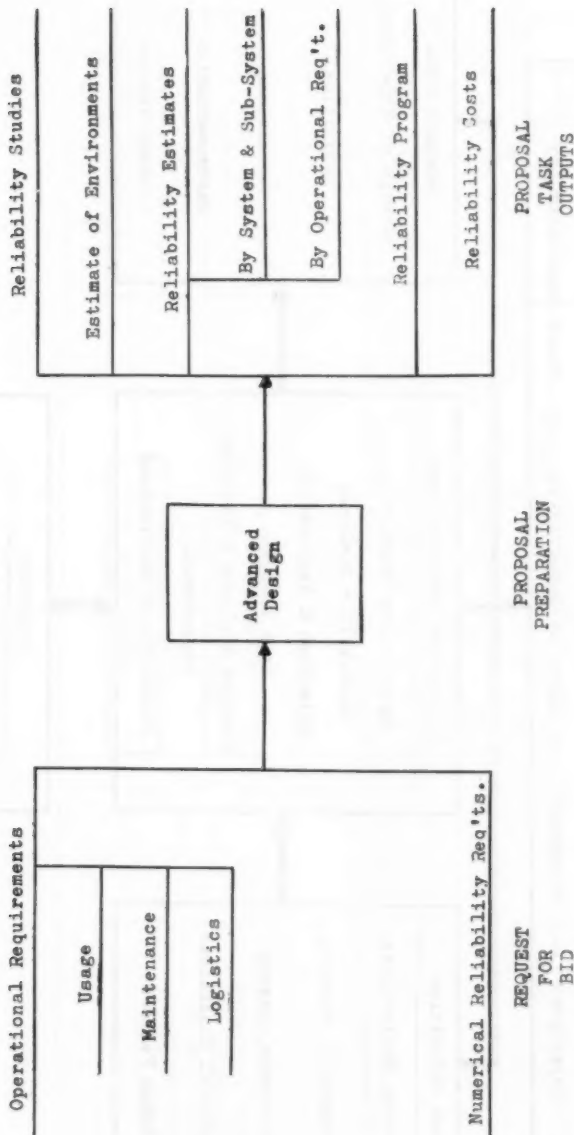
THE RELIABILITY PROGRAM
DESIGN & DEVELOPMENT PHASE

Organization

Objectives — Personnel — Lines of Supervision — Fixed Responsibility

Task Definition

Numerical Reliability Allowances

Environmental Criteria

Maintenance Concept

Operational Conditions

Human Factors

Typical Reliability Tasks

Reliability Prediction

Reliability Analysis

Selection & Application

Test Program

Failure Analysis & Feedback

Improvement

Knowledge of Environment

Demonstration

Design of Experiment

Definition

Conditions

Instrumentation

Significance

Controls

Trade-Off

Selection of Evaluation Points

Evaluation

Vendor and Subcontractor

FIG. 4

ORGANIZATION FOR RELIABILITY

THE MARTIN CO., Baltimore, Md.

FIG. 5

FIG. 6

594

SOLVING LINEAR PROGRAMMING PROBLEMS WITH CONTROL CHARTS

By

J. L. Dolby
General Electric Company

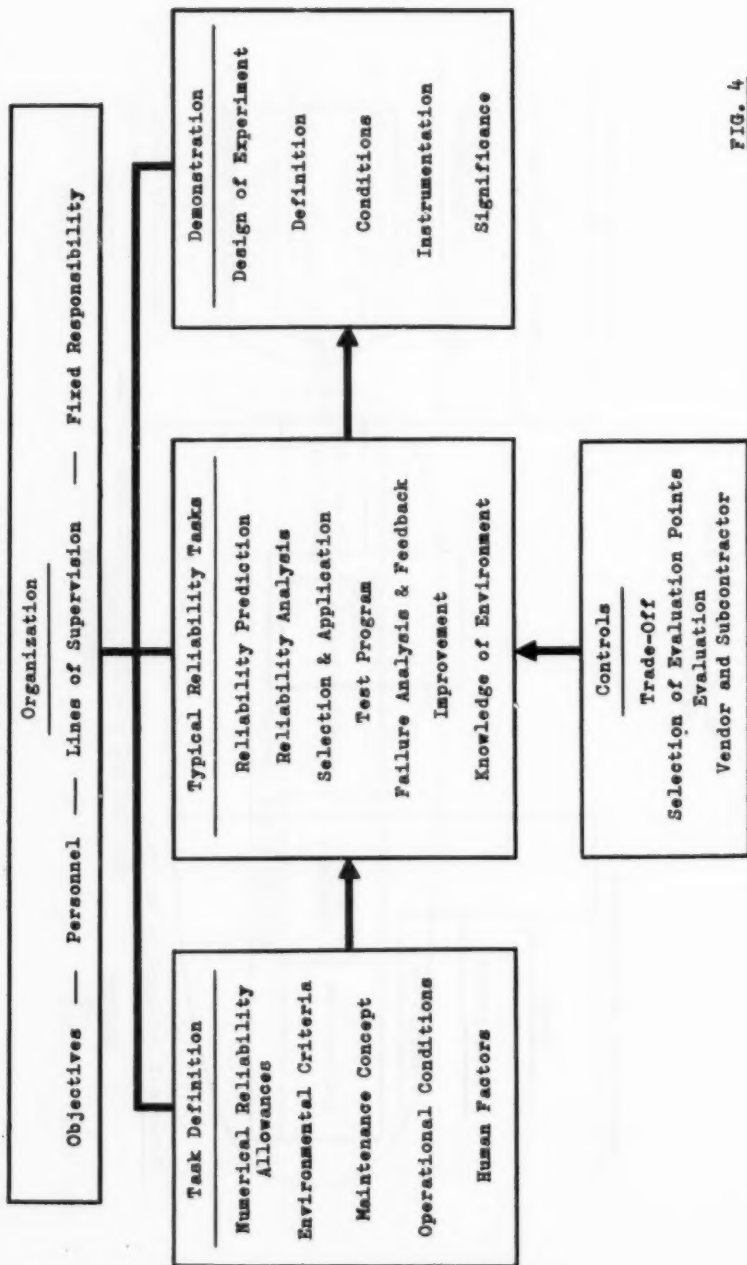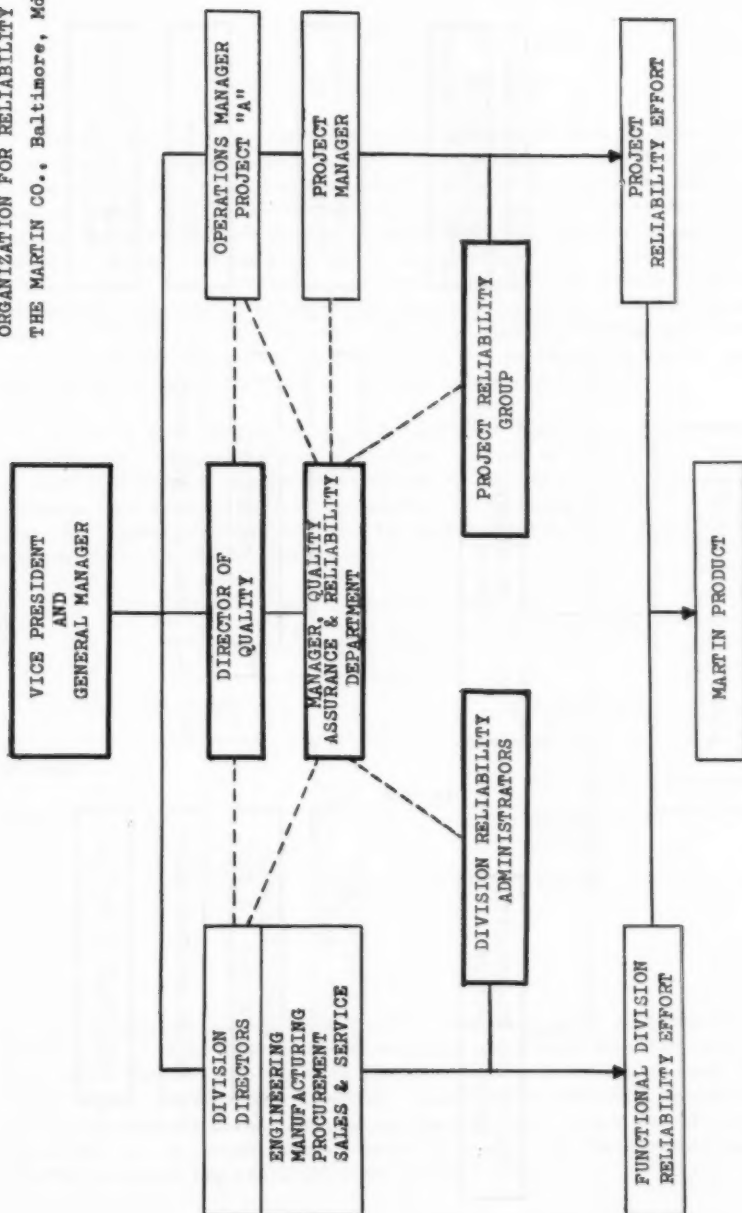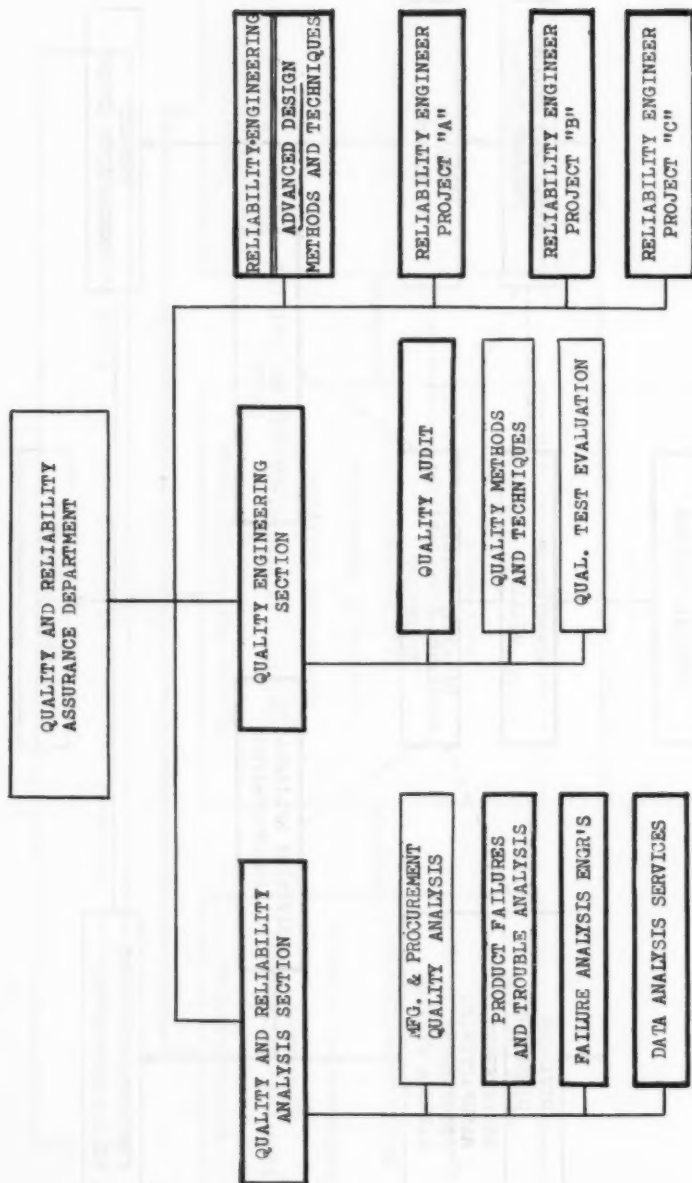In the past several years there has been a good deal of discussion
about linear programming as a handy tool for solving industrial problems.
On the other hand, there has been nearly as much discussion in the direc-
tion of: What is it? How do you use it? Who has used it successfully?
Outside the general area of what we might call "economic" problems, the
number of successful applications in industry has been rather limited.
At least part of this limitation has been caused by the rather formidable
mathematics involved in deriving the various solutions presently avail-
able. In this discussion we will present a somewhat simpler procedure
that is probably relatively inefficient in the mathematical sense, but is
much easier to understand and to explain to someone else.

First we must define the general problem. Let's take a hypothet-
ical example. Suppose we have a chemical process as in Figure 1, where
we wish to determine the proper   volume of the vessel, the operating
pressure, and Monomer to catalyst ratio. In so doing, we wish to mini-
mize the impurity content in PPM. We have available an equation that
relates these in the following way:

$$\text{Impurity content (PPM)} = 100 - 16 \text{ (volume)} - 4 \text{ (pressure)} + 12 \text{ (ratio)}$$
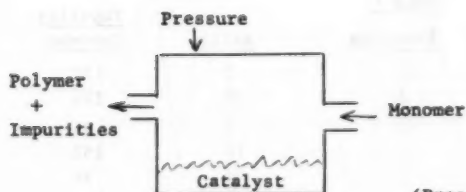
RESTRICTIONS:

$2 \leqslant \text{volume} \leqslant 12$ liters

$2 \leqslant \text{pressure} \leqslant 7$ atmospheres

$3 \leqslant \dfrac{\text{Monomer}}{\text{Catalyst}} \leqslant 10$

(Pressure) (Volume) < 15

Figure 1

It is the introduction of these restrictions that makes programming a
useful tool. There have been mathematical tools available for some time
that would enable us to find the maximum of many functions as long as
there were no restrictions involved. However, in realistic industrial
situations we must always face up to the fact that length cannot exceed
12 inches, or the weight cannot exceed 10 pounds, or the assembly must be
capable of operating efficiently at 120° F., etc.

Linear programming then appeals to everyone in its initial promise, precisely because it enables us to tackle realistic problems. It is, by definition, restricted to mathematical expressions that are linear, and this would probably explain why it is more useful in economic circumstances than in any other area, since it is possible to express many economic situations by linear equations.

The drag on the further application of linear programming has been the difficulty in understanding it. Most papers are either highly technical or highly mechanical, i.e., they discuss the problem in the context of the theory of convex sets, or they provide a set of "cook-book" steps without detailing the reasons for making these operations. Unfortunately this is inherent in the nature of the procedure. It is a very neat mathematical method, but its derivation is complicated and it is extremely difficult to give an intuitive description of it. However, if we are only interested in obtaining the minimum cost of a particular operation we can always use the trial and error procedure. In particular, we can make random trials, evaluate the cost of each trial, and select that combination that provides the lowest cost.

Going back to our example, suppose the volume must lie between 2 and 12 liters, the pressure between 2 and 7 atmospheres, and the catalyst ratio between 3 and 10. Furthermore, to meet temperature restrictions, the product of pressure and volume must not exceed 15. We could then select random numbers between 2 and 12 for the volume, 2 and 7 for the pressure, and 3 and 10 for the ratio, and calculate the impurity content. We would, of course, throw out any solutions that exceeded the temperature restriction. Twenty-five trials are listed below:

TABLE I

Trials	Volume	Pressure	Ratio	Impurity Content
1	3	4	7	120
2	2	3	10	176
3	4	3	6	96
4	3	5	10	152
5	4	2	4	76
6	2	4	6	124
7	3	4	5	96
8	5	2	7	96
9	2	5	5	108
10	4	3	4	72
11	3	2	8	140
12	3	2	3	80
13	4	2	5	88
14	6	2	7	80
15	2	5	8	144
16	4	3	7	108
17	3	2	9	152
18	3	2	4	92
19	4	2	7	112

TABLE I (CONT'D)

Trials	Volume	Pressure	Ratio	Impurity Content
20	6	2	5	56
21	2	4	10	172
22	5	3	8	104
23	3	2	4	92
24	3	5	8	128
25	7	2	9	88

Of course, this would not lead to the exact answer unless we took an infinite number of trials. However, as quality control people we do not normally require an exact estimate of the quality of a lot of material. We are perfectly willing to sample the lot because we know that beyond a certain point the information gained by further sampling does not equal the cost of further work. Even though this is a purely mathematical problem the same principle holds here. If we were doing this problem by hand, we would continue only until we were satisfied that further work would not sufficiently improve our answer to justify the cost.

As a means of determining when we should quit we can plot the results on a control chart, ignoring for the moment, the particular combination of length, width, and weight that we choose, and record only the impurity level. Figure 2 shows such a plot of the 25 trials listed above, with 3σ control limits.
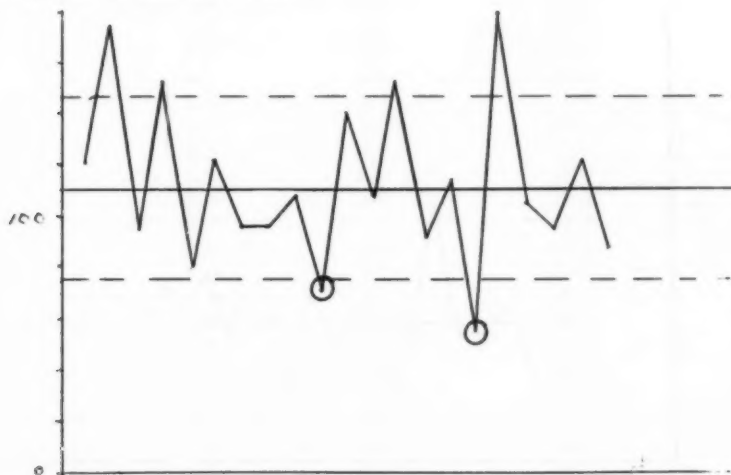

Figure 2

In this case, as with most control charts, we are going to be particularly interested in points that lie beyond the control limits. Since only two points fall below the limits in the first sample of 25, we took

597

another 50 trials to obtain a better idea of what conditions would lead
to points on the low side. At the end of these 75 trials we had 8 points
out of control low:

TABLE II

Volume	Pressure	Ratio	Impurity Content
4	3	4	72
6	2	5	56
5	2	3	48
6	2	3	32
7	2	5	40
6	2	3	32
4	2	3	64
7	2	5	40

On examination it turns out that the volume of the vessel is always
between 4 and 7 liters, the pressure is between 2 and 3 atmospheres, and
the catalyst ratio between 3 and 5.

Since in this case we want the low values of impurity content
we want to generate some more points that will fall in this low region
to give a better idea of low impurity performance requirements. We then
calculated 20 or more results using the new limits defined by the low
points of the initial runs. These results are summarized in Figure 3.
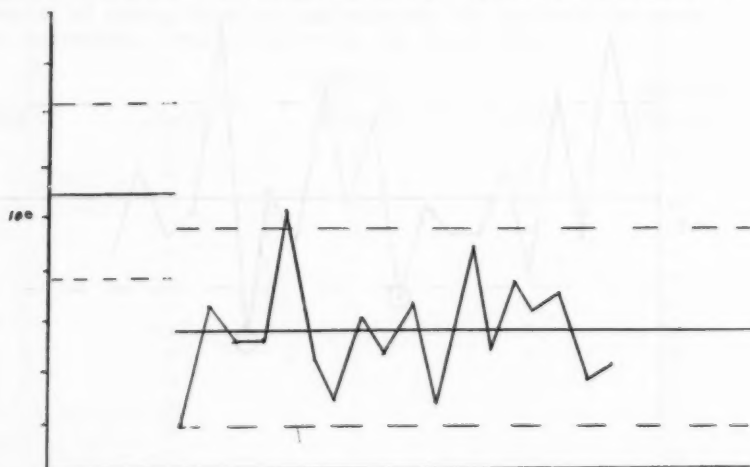

Figure 3

We see immediately that virtually all these trials are below the low
limit of the previous control chart. Furthermore, we find that the first

of these trials leads to an impurity content of only 16 PPM.

We could, of course, continue this analysis another step, but this should be enough for demonstration purposes. We can, however, get further useful information from this approach. Clearly, the volume of the vessel used in this hypothetical process would be a constant. However, we can expect that the pressure and the catalyst ratio will vary in actual operation. We could then set the volume constant (at 7 atmospheres since that was the value for the best run) and choose random values of the pressure and catalyst ratio as close to the optimum as is feasible. For this particular case we might say that pressure can vary between 2.0 and 2.1 and the ratio can vary from 3.0 to 4.0. Twenty five such runs are given in Table III. We can then expect that impurities will range between 15 and 26 PPM. A definitive statement would depend upon assumptions about the distribution. (We have used uniform distributions in all these examples.)

TABLE III

Pressure	Ratio	Impurity Content
2.0	3.8	25.6
2.1	3.6	22.8
2.0	3.1	17.2
2.1	3.7	24.0
2.1	3.8	25.2
2.0	3.8	25.6
2.0	3.1	17.2
2.0	3.6	23.2
2.1	3.1	16.8
2.0	3.5	22.0
2.0	3.2	18.4
2.1	3.0	15.6
2.1	3.3	19.2
2.0	3.0	16.0
2.0	3.3	19.6
2.0	3.1	17.2
2.0	3.7	24.4
2.1	3.8	25.2
2.1	3.2	18.0
2.0	3.4	20.8
2.1	3.6	22.8
2.1	3.7	24.0
2.0	3.1	17.2
2.1	3.2	18.0
2.1	3.3	19.2

In summary this particular procedure is quite simple to understand and to use. For very small problems its mathematical efficiency is poor, but even this difficulty is wiped out as the problems become more complex. Furthermore this procedure is not limited to linear problems,

nor does it require stringent mathematical assumptions. These basic
advantages should be sufficient to establish this as a working tool in
industry.

RESEARCH IN THE ORBIT OF QUALITY CONTROL

Paul E. Allen
Beech Aircraft Corporation

Research in the orbit of Quality Control has been going on for many years by experts far greater than I; however, the results of these efforts published in voluminous forms have left me confused as to their objectives and even greater still as to just what members of our profession think Quality really is -- is Quality making a part precisely match a drawing, or is Quality making a part that will fulfill its mission and provide customer satisfaction? I repeat -- is Quality making a part precisely match a drawing, or is Quality making a part that will fulfill its mission and provide customer satisfaction?

With a firm understanding of Beech Corporate Management philosophy of product quality, my own personal curiosity, and the cooperation of hundreds of you in Industry, I have tried to connect into an integrated program the best results of research in many of the apparent tangential concepts orbiting around the quality axis.

For the purpose of reader interest, I will start with our Corporate Management philosophies, our facilities, our products and our customers to give you a better understanding of why the efforts outlined in this paper actually have taken, or are taking, place.

Beech Corporate Officials have gone on written record to all BEECHCRAFT employees expressing their views of product quality, which briefly are:
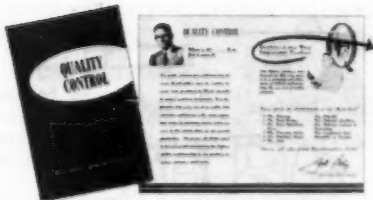
PRESIDENT AND CHAIRMAN OF THE BOARD

Quality, to me, is that invaluable characteristic that makes one brand better than another. A quality product is dependable . . . outstanding in appearance and performance . . . and competitive in price.

President

VICE PRESIDENT - GENERAL MANAGER AND DIRECTOR

QUALITY IS OUR MOST IMPORTANT PRODUCT

Vice President-General Manager

601

An Upward Sales Trend is Possible Only Through
SATISFIED CUSTOMERS

Fred E. Hedrich
Vice President-Coordinator

You can see from the above that we in Beech Quality Control have our objectives pretty well defined by our Corporate Management.

It was Walter Beech's wish that Beech have a conscience -- that conscience would be responsible for assuring that BEECHCRAFT products were of the best comparative quality and were able to fulfill their intended mission. In fulfilling Walter Beech's wish, Beech Management combined Quality Control and Customer Service into a single department and gave me the assignment of Department Manager. The reasoning behind this decision was to facilitate the implementation of an adequate data feedback system to assure a maximum degree of product evaluation which is necessary to properly control the product quality.

The Beech facilities are located in several factories, offices and test units in Wichita, Herington and Liberal, Kansas, and Boulder and Longmont, Colorado.

The Beech products consist of:

I. Military --

T-34   L-23   KDB-1 Target

MA-3 Power Unit   340 Analyzer   Dewer   Buddy Tank
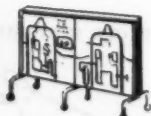
1700-Gallon Tank   Training Aids   Training Devices   Mobile Training Units
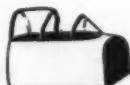
Technical Data

602

II. Commercial --

Bonanza   Travel Air   Twin-Bonanza   Super 18

III. Subcontract --

Wings   Aft Fuselages   Assemblies   Fuel Tanks

The Beech customer's list contains many Governments and elite of the free world.

Our products must satisfy the requirements of all applicable Government Agencies including CAA, Air Force, Army, Navy, Atomic Energy Commis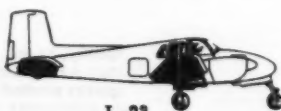sion, several prime weapons systems contractors, and about 800 annual commercial, domestic and export purchasers; therefore, our product control system must be versatile -- effective, yet not cumbersome.

Now with this background history, you can see that our Quality job is clearly defined and simple -- provide outstanding quality products that will fulfill their intended mission and provide customer satisfaction.

Up to this point, you can see that everything is clear-cut with no need for exploration or research -- now we explore the implementation of a plan to provide the results required by Management. All we have to do is:

    (1) Provide top product quality.
    (2) Provide product ability to fulfill its intended mission.
    (3) Provide customer satisfaction.

(I am sure you will all agree that you do not have a true quality product unless your product fulfills the above three basic prerequisites.)

One would normally assume that the task of implementing a product control program to fulfill three simple requirements, such as those listed above, could hardly necessitate any research effort -- these are pure cut-and-dried age old traditional philosophies and habits -- or are they?

When we find Manufacturing, Inspection, Procurement, Engineering, Service and Sales discussing a quality item under induced warm climatic

conditions sometimes augmented by hypertensive physical conditions, I must reconcile myself into asking, "What is Quality? Where is it defined?" The only fair way in my opinion to arrive at the answer to this basic question is to ask expert, novice, customer, designer, seller, builder, inspector and service men, "What in their opinion Quality really is", and arrive at a composite opinion.

Yes, this is the logical, simple approach to the problem and, once we have carried it out, we have eliminated the problem once and for all.

We make a questionnaire and ask 50 people the question, "What in your opinion is the job of Quality Control?" -- just as quick as we get 50 answers, we find that the job of Quality Control is to inspect parts to assure conformity to the engineering data.

Is this our cure-all answer? Well, the Customer Service Department says a large portion of our customer dissatisfaction comes from design items -- the customer says the Quality is poor, because it won't work.

Hmmm -- Inspection says it's like the drawing -- yet it won't fulfill its mission and the customer is unhappy. So our first research answer definition of Quality Control's function is inadequate as it does not appear to fulfill 2 of our 3 basic requirements.

Perhaps the 3 basic requirements are wrong -- so we ask over 100 people from 15 industries and Government Agencies:

(1) When you buy an article such as a car, boat, motor, radio, television set, etc., what do you expect from that article in the way of Quality?

    (a) Good workmanship?
    (b) Conformance to design?
    (c) Ability to run and do its intended job?

(2) How would you define Quality in one sentence?

(3) What in your opinion is Quality Control?

The answers to this inquiry, which are shown on Figure 1, average out as follows:

(a) 66% of the people considered ability of a product to do its intended job was the most important quality factor.

(b) 23% of the people considered good workmanship was the most important quality factor.

(c) 11% of the people considered conformance to design the most important quality factor.

Quality Control was generally defined as being: The technique, program or policy used in attaining and assuring the quality level.

Quality was generally defined as being: The measure of the product or service relative to its ability to perform and provide customer satisfaction against some standard of performance of perfection.

Figure 1.

Well, the recap of this study would appear to indicate it is general opinion that a quality product should be able to fulfill its mission, have good appearance and supply adequate customer satisfaction. So our first 3 basic requirements for Quality Control still appear to be applicable.

Maybe Customer Service is wrong and design isn't a major factor in product quality -- so we ask Management or Command personnel serving in Engineering, Quality Control, Manufacturing, Testing and Flight Safety Groups in the aircraft, missile, engine and equipment industry, and the Air Force, Army, Navy, Airlines and the CAA, the following questions:

What percent of a product's service problems would you estimate are due to:

(a) Design?
(b) Nonconformance to design?
(c) Vendor component not suited to a system?
(d) Misuse of the product by the customer?
(e) Other?

The results of this research questionnaire which are shown in Figure 2 reveal that, in the opinion of those polled, the customers' service problems are:

(a) 38.6% due to design.
(b) 25.8% due to misuse of the product.
*(c) 12.6% due to vendor component not suited to the system application.

605

(d) 11.2% due to nonconformance to the design.

(e) 11.7% due to other causes.

*In most cases, these were specified by design.

	PRODUCT CONTROL AND RELIABILITY QUESTIONNAIRE SUMMARY SHEET AVERAGE					
	CATEGORY OF WORK	ENGINEERING AVERAGE	QUALITY CONTROL AVERAGE	SERVICE AVERAGE	MILITARY USER AVERAGE	GENERAL AVERAGE
PERCENT PRODUCTS SERVICE PROBLEMS DUE TO:	DESIGN	50%	41%	27%	30%	38.6%
	NONCONFORMANCE TO DESIGN	11%	13%	8%	10%	11.2%
	VENDOR COMPONENT NOT SUITED TO SYSTEM	4%	17%	12%	15%	12.6%
	MISUSE OF PRODUCT BY CUSTOMER	26%	20%	33%	30%	25.8%
	OTHER	7%	9%	20%	15%	11.7%

Figure 2.

(It is interesting to note that Engineers felt 50% of the problems
were due to design while the mixed average was only 38.6%, and the
Inspection personnel felt 13% of the problems were due to nonconformance
to design while the mixed average was 11.2%. The military users felt
that 30% of the problems were due to misuse while the mixed average was
25.8% -- it appears that each Group imposed more self-criticism than the
over-all average.)

Now we have generally defined Quality and we have generally located
where the areas are where Quality Control is needed -- how do we get it?
This should be the easiest task of all as our Library has many books and
periodicals pertaining to the subject, so all we need to do is read a
few stories and we are in business managing a Quality Control operation.

Well, at least it should be simple -- or shouldn't it? We review
many issues of our ASQC's periodical and find that about 76% of the
articles pertain to the scientific expose of the mathematical applica-
tion of numerical quantities of variance or lack of variance of a char-
acteristic of parts from a specification or drawing. Our researcher
failed to report any significant amount of text devoted to areas of
Quality Control other than conformance to design wherein only 11.2% of
the customer problems exist.

Hmmm -- maybe the simple task of Quality Control implementation is
just complex in the one area, nonconformance to design, where the 11.2%
of the customers' problems are, as most of the ASQC periodical effort
is, in this area -- should we assume this to be true and go on. Well,
we're skeptical, but we have to assume something, so we go along with
the statistical theorist -- to a degree -- and find that product quality
as pertaining to design conformance can be better guided by application
of statistical Quality Control to the manufacturing process by charting
the results of sample inspections.

This makes good sound logic until our researcher -- a precise,
persistent, persuasive personality comes in from the Library with a
report from a well-known company with quotes from Universities, etc.,

which says that 100% inspection will only catch 68% of the defective
parts and 400% inspection will only catch 98% of the defective parts.

Well, with this revelation, you realize that it's a good thing
that only 11.2% of the customers' problems are due to nonconformance
to design and proceed to put your Quality Control Research Engineer to
work by ordering a research study to explore a few simple basic prin-
ciples of product control, such as:

(a) The best type of human being to inspect various categories
     of parts.
(b) The cost of inspection vs. accuracy of inspection by
     different methods.
(c) The hypnotic effect of repetitive inspection on different
     types of people.
(d) The comparative inspection ability to catch the defects
     in a lot of parts when the same lot is checked by the
     inspection system in five different companies.
(e) The ability of Engineers to check parts to their own
     designs.

This program has just started; however, the preliminary results
are indeed interesting, such as:

(1) Although the pay rates of the Engineers selected for the
     tests averaged 2.2 times that of the Inspectors, their
     performance was 8.4% less effective.
(2) The cost of equipment necessary to check the parts, as
     determined by two Inspectors, varied from $858.20 to
     $3,106.75.
(3) The man with the lowest inspection equipment cost re-
     quirement was 98% effective, while the man with the
     highest equipment cost requirement was 98% effective
     in catching the defective parts.

Another area of research is the vendor-furnished item failures and
the causes for these failures. In reviewing the area of vendor items
that fail to function in a system, we find several factors contributing
to the problems, such as:

(1) The test article sent in by the vendor is a top, rather
     than average, performer.
(2) Some vendors seem to be more interested in getting the
     order than they are in the suitability of the component
     for the job.
(3) Some suppliers appear to have little interest in their
     product or reputation as pertaining to quality or product
     reliability.

The results of these preliminary research efforts indicate that
the term Quality Control cannot be limited to the checking or inspec-
tion of materials or parts for conformity to a design, but must assume
a much broader scope if our intent is to provide products that have the
functional quality required to fulfill the intended mission.

Possibly we in ASQC should evaluate the distribution of effort by
our Society so we publicize and discuss with proper proportions each

area of product control and not lean specifically to pure statistics.

Possibly we should more fully evaluate human capabilities so as to validate the statistical assumptions.

But most of all, let us not forget that our whole objective should be to assure customer satisfaction -- without customers for our products, our future would be very limited. May the Good Lord never let us forget that good customers are hard to get.

QUALITY'S NEGLECTED CUSTOMER - THE ENGINEERING DEPARTMENT

Henry A. Schumer
Chief Engineer, Asheville Plant
International Resistance Company

If the Quality Control Department is to make a significant con-
tribution, it must recognize the Engineering Department as a customer
and it must recognize the customer's needs just as it does in the usual
customer - vendor relationship. Thus, the Quality Department must
evaluate its relationship with Engineering by asking itself these
questions:

1. Is it providing a service which satisfys Engineering
    needs?
2. Is this service more effective than others available?
3. Is an acceptable product delivered, on time?

Unless the answers to all of these questions are yes, the use of
statistical quality control techniques may result in a less effective
solution to a problem than if some other more orthodox (from the
Engineering Department point of view) technique was used.

It is not unusual for the Quality Department to be in a position of
determining whether the product which it has to sell is consistent with
the customer's needs. Its usual function is to evaluate the product of
others to establish whether it is in truth satisfactory. However, the
same concept should apply when dealing with Engineering problems, except
in this case it is one of soul searching to determine whether the tech-
niques which are available to the Quality Engineer are necessary for
performing the job at hand, or whether there are other methodologies
which would provide a cheaper, faster, but equally as efficient or
effective answer. In many areas the most efficient technique may be
nothing more than "practical engineering judgement"; whereas, the added
sophistication and cost of a statistical treatment of the problem may
be the most inefficient. It is, therefore, our obligation to recognize
and assist the customer in deciding which techniques are applicable. In
most instances, the measurement or success of a solution is determined
by the cost savings, reliability, or conformance to specification limits.
A careful evaluation of the consequences of a wrong decision must be
made and unless it is clear that the time invested in the use of some
subtle technique can be justified, advocating its use would be improper.

Once we have established that the problem is one which is within
the realm of statistical quality control, and by this we mean economi-
cally and effectively within the realm of statistical techniques, it is
important that we dispel much of the mystery involved in the use of this
science. We, the practitioners of the art, must not only be able to
understand the limitations of statistical techniques, but also be capa-
ble of putting them in their proper perspective so that we do not fall
into the trap of making a decision which, while technically sound, many
times has little meaning in the business or operational sense.

The third factor which must be considered is the delivering of an
acceptable product on time, and it is here that many quality investi-
gators are guilty of unclear thinking. The acceptable product which
must be provided is a complete analysis of the data, so that the "line"

609

function (the Engineering Department) can make the correct decision, and not a decision as to what future action is required. On the surface this may not seem to be a significant problem; however, unless this principle is closely adhered to, the relationship between Quality and Engineering will be fraught with misunderstandings and mistrust and as a result could very well negate the long term value of the Quality Department. As the Quality Control - Engineering relationship matures, suggestions and recommendations presented by Quality will not only be expected but will be sought.

The maturing process requires the complete removal of suspicion regarding methods or techniques as well as intent, and since suspicion of new things is a natural tendency among most people it is best handled by familiarization with the new technique or method. Wherever possible, the Engineering staff should be provided with "Quick and Dirty" techniques of a "Cook Book" type so that they may become familiar with the methods involved. A further extension of this technique would be in the form of a training agenda which might include the following topics:

1. Fundamental statistical techniques as applied to Engineering problems.
2. Introduction to experimental design.
3. Review of previous Engineering decisions based on the use of these techniques.

The company organization and position of the Quality Department is also a critical factor in determining the success of Quality Control techniques in the Engineering function. However, regardless of its area of responsibility, in the final analysis the success of the program will be directly related to the sales ability of the practitioner. In my experience I have been actually responsible for the failure of a program by forcing of issues because of company position and as a result alienated the persons involved so as to reduce the overall effectiveness. A sound foundation for this activity must be based on mutual trust, confidence, and understanding, and cannot be ordered in. Therefore, the program must be carefully planned and thoroughly sold beforehand.

It is true that individual battles can be won and in some cases recalcitrants must be shown the way; however, it is obvious that this method of conversion is to be avoided when possible. The best approach to sell this program is to convince the Engineer customer that he needs what you have to sell and that you can make his job easier or make him more efficient. In a short time this should lend to a clearer awareness of the general field of application and perhaps to requests for application of more sophisticated techniques on Engineering data. Have no fears - Engineering personnel are too occupied with engineering problems to become expert statistical analysts (even Quality Engineers do not enjoy this luxury).

Unfortunately, only the minor effects of Quality Control assistance to Engineering are recognisable in dollars and cents; a reduction in experimental time, the reduction in test samples, replication of experiments, etc. The really significant gains in increased efficiency of decisions and reduction of overall development time from conception to factory operation are not readily evaluated but are unquestionably the major factors.

Over the years IRC recognized the need for "a procedure to delegate responsibility for initiating, developing, and manufacturing of new products". In 1953 Mr. G. Entrekin, then Chief Product Engineer, and Mr. L. Jacobson, then Manager of Quality Control, developed the following procedure which has resulted in many of the benefits discussed above.

Table I illustrates the mechanism of the procedure whose salient points are:

1. Assignment of project by top management to a project leader (usually Research Division) whose overall responsibilities are clearly defined.
2. The review of the project by top management after preliminary evaluations have been made.
3. If in the opinion of top management the market potential, product performance, and profit potential dictate continuance of the project, the project is turned over to Engineering for further development.
4. After the Engineering development is complete, which includes design of processing equipment, all tool and product drawings, manufacturing procedures, quality standards, cost standards, etc., a pilot operation is set up jointly by the operating plant personnel and the engineering group with all plant staff services participating in both the design of the pilot plant operation as well as evaluation of the results.

Following this system materially reduces the time and money involved in the development of a new product. It also provides all groups and departments with an opportunity to review and comment on a product prior to its introduction to production with the concomitant confusion and delay which results from attempting to do engineering on the production floor.

1. ITEMS OF RESEARCH ORIGIN (PHASE A)

A. Assignment of a Project

TABLE I
NEW PRODUCT & PROCESS RELEASE PROCEDURE

Organizational Responsibility	Responsibilities (Check-List)	Suggested Staff Services
Executive Committee	1. Decide need for Research project and assign to Director of Research along with objectives to be met.	

B. Research Responsibility Stage

Research Division	1. Do initial creative work.	
	2. Make experimental models of proposed design. ("Proposed Design" as used herein includes both physical designs and proposed processes).	
	3. Conduct performance tests of proposed design.	
	4. Review with Engineering and Quality Control, process considerations.	
	5. In connection with Engineering and Quality Control, design an experiment and make a suitable quantity of units of the proposed design to evaluate the proposed product and process.	Engineering representatives to be designated by Chief Engineer.
	6. When it is mutually agreed that the objectives of the project have been met, the Director of Research will submit a report on the project to the Executive Committee.	

Organizational Resp.	Responsibilities	Staff Services

Executive Committee	1. Decide if the project should be carried further by Engineering. If so, assign the project to Engineering along with:
	
	  a. Type, size and performance characteristics required of the product.
	  b. Estimated yearly requirements.
	  c. Maximum breakeven costs.	

Chief Engineer	1. Assign Phase A portion of project to one of Engineering Departments.	
C. Engineering Responsibility Stage		
Responsible Engineering Department	1. If necessary, review with the Research Division the product and process as developed, and prepare a time schedule for all Phase A work.	Research Division
	2. Prepare initial ES drawings of parts, and review, with the staff services shown, for dimensional tolerances and other requirements.	Drafting Section Product Engineering Quality Control Mechanical Engineering
	3. Review, with the staff services shown, process considerations.	Methods & Stds., Product Engineering, Mfg. Depts., Mechanical Engineering, Quality Control, Product Engineering, Mfg. Departments Purchasing
	4. Estimate shop losses and yields.	
	5. Consider availability of materials.	
	6. Consider scheduling aspects of manufacture.	Production Control
	7. Run a pilot run of models.	Same as items 1-6 plus Mfg. Depts. for pilot run.

613

Organizational Resp.	Responsibilities	Staff Services
	8. Review design in light of items 2-7 inclusive and modify where necessary.	Same as items 1-6
	9. Obtain breakeven cost estimates.	Accounting
	10. Obtain preliminary Sales Review and approval.	Sales
	11. Obtain preliminary estimate of tooling costs.	
	12. Prepare general schedule of Phase B activities.	Mechanical Engineering
	13. Prepare report to Chief Engineer (copy to Chief Product Engineer and Director of Research) of all Phase A results.	All interested departments

2. ITEMS OF RESEARCH OR ENGINEERING ORIGIN (PHASE B)

A. Process Development

Organizational Resp.	Responsibilities	Staff Services
Responsible Engineering Department	1. Assume the overall responsibility for all Phase B activities, utilizing where necessary the staff services of any and all departments of the company.	All Departments including Research when Phase A has been done by them.
	2. Call a meeting of representatives of all interested departments and after reviewing the product design and recommended manufacturing process, review the general schedule previously approved for this phase of activity. Prepare a tentative detailed schedule of activities. List all items which require action, the types of action required, and the department and person whom it is recommended will be directly responsible for each action.	

614

Organizational Resp.	Responsibilities	Staff Services
	3. Request the development, design and detailing of all necessary manufacturing equipment.	Mechanical Engineering assisted when requested by Methods and Stds. Product Eng., etc.
	4. Request the design of all necessary gages.	Manufacturing Engineering, Prod. Engineering, Quality Control
	5. In conjunction with Mechanical Engr., determine the department responsible for the procurement, installation and manufacturing of each item in 4. and 5. above. See that necessary paperwork is issued.	Mechanical Engineering
	6. Arrange for all space, safety and utility needed.	Industrial Engineering
	7. Complete development work where necessary on materials and prepare tentative material specifications.	Product Engineering Chemical Engineering
	8. Freeze design and request preparation of final ES (experimental sketch) drawings for all parts.	Drafting
	9. Prepare tentative quality and inspection standards.	Product Engineering Quality Control, Sales
	10. Obtain sampling and inspection plans for incoming materials, and product manufacture.	Quality Control Purchasing
	11. Review procurement and manufacturing problems. Obtain estimated delivery cycles. Develop approved sources of supply.	Production Control Purchasing, Manufacturing Depts.
	12. Order all parts necessary for large scale pilot run.	Purchasing, Production Control
	13. Prepare tentative manufacturing procedures.	Product Engineering
	14. Prepare tentative job descriptions.	Methods & Standards, Industrial Relations, Manufacturing Dept.

Organizational Resp.	Responsibilities	Staff Services
	15. Prepare tentative technical information for Sales release.	Product Engineering, Sales, Quality Control
	16. Review scheduling and ordering procedure.	Production Control, Sales
	17. Call a meeting of the representatives of the departments listed in the right hand column and present to them the proposed ES drawings, material specifications, manufacturing procedures, etc. Jointly design a pilot run which will supply all interested departments with that information which they consider essential. This is to include a determination of tool and machine performance, adequacy of mfg. procedures, types of mfg. problems, yields, product performance, etc.	Product Engineering Chemical Engineering Mechanical Engineering Quality Control Methods & Standards Manufacturing Production Control Accounting Research when Phase A was done by them.
	18. Schedule pilot runs as a part of manufacturing schedule.	Production Control
	19. Obtain a recheck of product breakeven costs based on pilot run experience.	Accounting
	20. Obtain performance characteristics of product made in pilot run by production methods.	Product Engineering
	21. Request acceptance of product by Product Engineering for limited manufacture on the basis of the proposed drawings and specifications, and the result of the pilot run.	Product Engineering
	22. Request acceptance of product by Works for limited manufacture on the basis of the proposed drawings and specifications and the results of the pilot run.	Quality Control Manufacturing

Organizational Resp.	Responsibilities	Staff Services
	23. Determine a limited production schedule which is acceptable to the three staff services shown. At the same time decide the method of handling the product produced during this limited production - sell, hold, etc.	Production Control Product Engineering Sales
	24. Arrange for the procurement, expediting and storing of all parts necessary for limited production.	Purchasing Production Control
	25. Operated under standard factory conditions to produce limited quantities of product. (Use tentative specifications and mfg. procedures).	Product Engineering, Production Control, Quality Control, Mechanical Engr., Chemical Engr., Methods & Standards
	26. Eliminate all manufacturing difficulties and revise drawings and specifications where necessary.	Same as for 25
	27. Provide Works with listing of all jobs ready for rating.	Methods & Standards

B. Acceptance for Production

Organizational Resp.	Responsibilities	Staff Services
Product Engineering	1. Accept product for full scale production.	
	2. Have numbered and released all part and assembly drawings, mfg. procedures and material specifications.	Drafting
	3. Release to Works for full scale prod.	
	4. Release technical information for Sales Bulletin to Sales.	
Works	1. Accept product for full scale prod.	
	2. Rate all jobs.	
	3. Advise Sales of mfg. capacities and date of initial full scale prod.	
Sales	1. Release Sales Bulletin.	